Uptime Institute INTELLIGENCE

PLANNING & STRATEGY

UI Intelligence report 32

Ten data center industry trends in 2020

Industry dynamics, market developments, innovations and challenges

Lead Authors
Rhonda Ascierto, Vice President of Research, Uptime Institute
Andy Lawrence, Executive Director of Research, Uptime Institute

Contributing Authors
Rabih Bashroush, Research Director, Uptime Institute
Kevin Heslin, Chief Editor, Uptime Institute

Uptime Institute's examination of some of the top trends in data centers in 2020 shows an industry that that is confidently expanding, especially at the edge, attracting new investors and increasingly embracing new technologies. But it also faces some challenges — particularly in the areas of resiliency, staffing, environmental impact and energy use.

30-45 minutes to read

UII-32 v1.1 published October 25, 2019 last updated December 3, 2019

This Uptime Institute Intelligence report includes:

Table continues next page

ABOUT UPTIME INSTITUTE INTELLIGENCE

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.

ABOUT UPTIME INSTITUTE

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 and 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 US, Mexico, Costa Rica, Brazil, UK, Spain, UAE, Russia, Taiwan, Singapore and Malaysia. Visit 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

Introduction

The digital infrastructure industry continues to grow and change at a striking price. Across the world, a thriving community of investors, designers, owners and operators are grappling with many of the same issues: resiliency and risk, the impact of cloud, the move to the edge, rapid innovation, and unpredictable (although mostly upward) demand.

What should stakeholders in this industry expect in 2020? Which innovations will make a difference — and which have been exaggerated? Is the challenge of distributed resiliency being solved, or is it getting worse? What are regulators likely to do?

The ten trends the Uptime Institute Intelligence team has identified show an industry that is confidently expanding toward the edge, that is attractive to many new investors, and that is embracing new technologies and architectures — but one that is also running against some headwinds. Resiliency concerns, premature expectations about the impact of 5G, climate change, environmental impact and increasingly watchful regulators are among the hazards that must be successfully navigated.

TEN DATA CENTER INDUSTRY TRENDS IN 2020

1. **Outages drive authorities and businesses to act**

   Big IT outages are occurring with growing regularity, many with severe consequences. Executives, industry authorities and governments alike are responding with more rules, calls for more transparency and a more formal approach to end-to-end, holistic resiliency.

2. **The internet tilts toward the edge**

   In the coming years, significant data will be generated by many more things and much more will be processed away from the core, especially in regional data centers. Many different types of data centers and networking approaches will be needed.

3. **Data center energy use goes up and up**

   Energy use by data centers and IT will continue to rise, putting pressure on energy infrastructure and raising questions about carbon emissions. The drivers for more energy use are simply too great to be offset by efficiency gains.

4. **Capital inflow boosts the data center market**

   Data centers are no longer a niche or exotic investment among mainstream institutional buyers, which are swarming to the sector. New types of capital investors, with deep pockets and long return timelines, could boost the sector overall.

Continues next page

TEN DATA CENTER INDUSTRY TRENDS IN 2020 *(continued)*

5. **More data, more automated data centers**

   Many managers are wary of handing key decisions and operations to machines or outside programmers. But recent advances, including the broad adoption of data center infrastructure management systems and the introduction of artificial intelligence-driven cloud services, have made this much more likely. The case for more automation will become increasingly compelling.

6. **Data centers without generators: More pilots, more deployments**

   Most big data centers cannot contemplate operating without generators, but there is a strong drive to do so. Technological alternatives are improving, and the number of good use cases is proliferating. The next 24 months are likely to yield more pilots and deployments.

7. **Pay-as-you-go model spreads to critical components**

   As enterprises continue to move from a focus on capital expenditures to operating expenditures, more critical infrastructure services and components — from backup energy and software to data center capacity — will be consumed on a pay-as-you-go, "as a service" basis.

8. **Micro data centers: An explosion in demand, in slow motion**

   The surge in demand for micro data centers will be real, and it will be strong — but it will take time to arrive in force. Many of the economic and technical drivers are not yet mature; and 5G, one of the key underlying catalysts, is in its infancy. Demand is likely to grow faster from 2022.

9. **Staffing shortages are systemic and worsening**

   The data center sector's staffing problem is systemic and long term, and employers will continue to struggle with talent shortages and growing recruitment costs. To solve the crisis, more investment will be needed from industry and educators.

10. **Climate change spurs data center regulations**

    Climate change awareness is growing, and attitudes are hardening. Although major industry players are acting, legislators, lobbyists and the public are pressing for more. More regulations are on the way, addressing energy efficiency, renewable energy and waste reduction.

TREND ONE

Outages drive authorities and businesses to act

Big IT outages are occurring with growing regularity, many with severe consequences. Executives, industry authorities and governments alike are responding with more rules, calls for more transparency and a more formal approach to end-to-end, holistic resiliency.

IT outages and data center downtime can cause huge disruption. That is hardly news: veterans with long memories can remember severe IT problems caused by power outages, for example, back in the early 1990s.

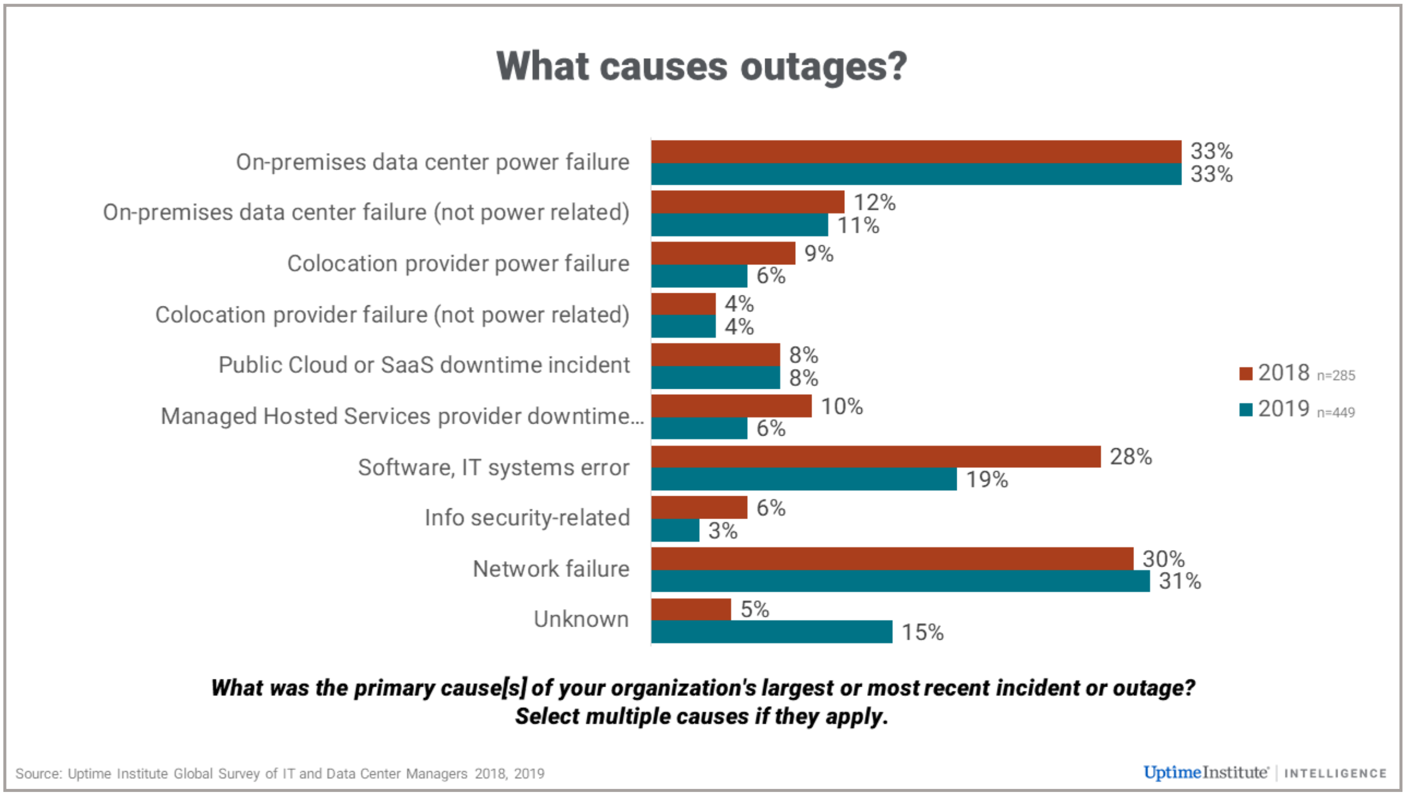
Three decades on, the situation is vastly different. Almost every component and process in the entire IT supply chain has been engineered, re-engineered and architected for the better, with availability a prime design criterion. Failure avoidance and management, business continuity and data center resiliency has become a discipline, informed by proven approaches and supported by real-time data and a vast array of tools and systems.

But there is a paradox: The very success of IT, and of remotely delivered services, has created a critical dependency on IT in almost every business and for almost every business process. This dependency has radically increased in recent years. Many more outages — and there are more of them — have a more immediate, wider and bigger impact than in the past.

A particular issue that has affected many high-profile organizations, especially in industries such as air transport, finance and retail, is “asymmetric criticality” or “creeping criticality.” This refers to a situation in which the infrastructure and processes have not been upgraded or updated to reflect the growing criticality of the applications or business processes they support. Some of the infrastructure has a 15-year life cycle, a timeframe out of sync with the far faster pace of innovation and change in the IT market.

While the level of dependency on IT is growing, another big set of changes is still only partway through: the move to cloud and distributed IT architectures (which may or may not involve the public cloud). Cloud and distributed applications enable the move, in part or whole, to a more distributed approach to resiliency. This approach involves replicating data across availability zones (regional clusters of three or more data centers) and using a variety of software tools and approaches, distributed databases, decentralized traffic and workload management, data replication and disaster recovery as a service.

These approaches can be highly effective but bring two challenges. First are complexity and cost — these architectures are difficult to set up, manage and configure, even for a customer with no direct responsibility for the infrastructure (Uptime Institute data suggests that difficulties with IT and software contribute to ever more outages). And second, for most customers, is a loss of control, visibility and accountability. This loss of visibility is now troubling regulators, especially in the financial services sector, which now plan to exercise more oversight in the United States (US), Europe, the United Kingdom (UK) and elsewhere.



Will outages get worse?

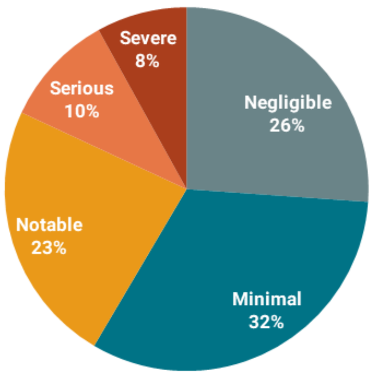
Are outages becoming more common or more damaging? The answer depends on the exact phrasing of the question: neither the number nor the severity of outages is increasing as a proportion of the level of IT services being deployed — in fact, reliability and availability is probably increasing, albeit perhaps not significantly.

But the absolute number of outages is clearly increasing. In both our 2018 and 2019 global annual surveys, half (almost exactly 50%) said their organization had a serious data center or IT outage in the past three years – and it is known that the number of data centers has risen significantly during this time. Our data also shows the impact of these outages is serious or severe in almost 20% of cases, with many industry sectors, including public cloud and colocation, suffering problems.

What next?

The industry is now at an inflection point; whatever the overall rate of outages, the impact of outages at all levels has become more public, has more consequential effects, and is therefore more costly. This trend will continue for several years, as networks, IT and cloud services take time to mature and evolve to meet the heavy availability demands put upon them. More high-profile outages can be expected, and more sectors and governments will start examining the nature of critical infrastructure.

One in five outages is rated Serious or Severe



Category 1 Negligible	Recordable outage, but little or no obvious impact on services.
Category 2 Minimal	Services disrupted. Minimal effect on users/customers/reputation.
Category 3 Notable	Customer/user service disruptions, mostly of limited scope, duration or effect. Minimal or no financial effect. Some reputational or compliance impact(s).
Category 4 Serious	Disruption of service and/or operation. Ramifications include some financial losses, compliance breaches, reputation damages, possibly safety concerns. Customer losses possible.
Category 5 Severe	Major and damaging disruption of services and/or operations with ramifications including large financial losses, possible safety issues, compliance breaches, customer losses, reputational damage.

Thinking of your most recent service outage, how would you categorize it? Choose one.

Source: Uptime Institute Global Survey of IT and Data Center Managers 2019, n=448 — UptimeInstitute | INTELLIGENCE

This is already started in earnest: In the UK, the Bank of England is investigating large banks' reliance on cloud as part of a broader risk-reduction initiative for financial digital services. The European Banking Authority specifically states that an outsourcer/cloud operator must allow site inspections of data centers. And in the US, the Federal Reserve has conducted a formal examination of at least one Amazon Web Services (AWS) data center, in Virginia, with a focus on its infrastructure resiliency and backup systems. More site visits are expected.

Authorities in the Netherlands, Sweden and the US have also been examining the resiliency of 911 services after a series of failures. And in the US, the General Accounting Office published an analysis to determine what could be done about the impact and frequency of IT outages at airlines. Meanwhile, data centers themselves will continue to be the most resilient and mature component (and with Uptime Institute certification, can be shown to be designed and operated for resiliency). There are very few signs that any sector of the market (enterprise, colocation or cloud) plans on downgrading physical infrastructure redundancy.

> As a result of the high impact of outages, a much greater focus on resiliency can be expected, with best practices and management, investment, technical architectures, transparency and reporting, and legal responsibility all under discussion.

TREND TWO

The internet tilts toward the edge

In the coming years, significant data will be generated by many more things, and much more will be processed away from the core, including in regional data centers. Many different types of data centers and networking approaches will be needed — and this is driving more innovation and new types of infrastructure.

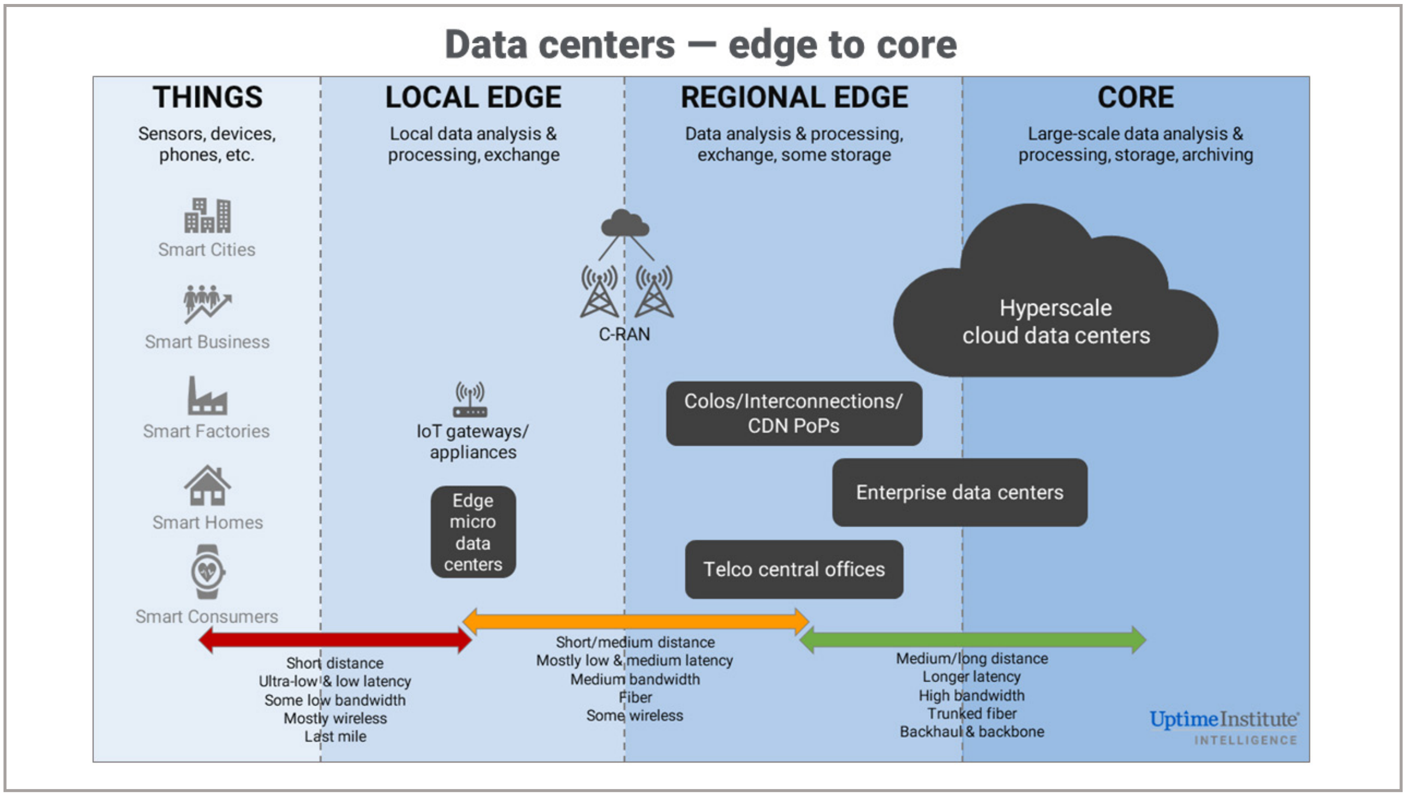
Edge IT is growing across the board, from mobile devices and streaming content to newer workloads such as the internet of things (IoT), multi-player online gaming and augmented reality. In many cases, it will be cheaper or faster to process this data close to where it is generated or used. Yet today's dominant IT model is centralized, with most data traversing a wide area network to and from a data center sited farther than 50 miles away. IT infrastructure — and the internet — has grown much, much bigger but not closer, or at least not close enough to support the anticipated wave of edge computing.

A rising tide

Edge computing involves many players: cable companies, mobile network operators, cellular tower companies, content distribution network (CDN) providers, data center owners, governments, landowners, service providers, software companies and more. How these companies will coordinate resources and intelligence between nodes, how they will interface with legacy networks, and how they will profit — the answers to all these questions are works in progress. Overall, ownership and management of edge capacity will be diverse and will depend on the function and specific applications.

Today's providers of large-scale data center and network capacity will clearly play a major role. Many (if not most) large and medium-sized data centers such as colocation facilities are already sited deliberately or fortuitously near where most edge data is either generated or consumed, at the regional edge. In terms of latency, bandwidth and connectivity, the regional edge is typically in the next "zone" from the local edge, where sub-millisecond or low-single-digit latency is required (see graphic). In spite of the anticipated demand for micro data centers at the local edge (see **Micro data centers: An explosion in demand, in slow motion**), the vast majority of edge applications — some say 95% — will not require the very low level of latency that the local edge delivers and could be probably be supported by data centers at the regional level.

The following illustration presents a broad, high-level schema of the edge data center ecosystem.



By proxy of their location and proliferation, colocation providers will play a key role, at the regional edge. Colo data centers offer near-local storage, analysis and integration of edge data and, increasingly, interconnections for network peering at the edge. Many are positioning their interconnected facilities as having cost and performance benefits over centralized enterprise and public cloud models. More are offering virtual interconnects and combining interconnection and colo with edge services, such as tiered cloud-integrated storage. And they are building and buying capacity in new regions to expand their footprints in the regional edge.

Colos are also home to CDN providers, which are also set to play a big part in the expansion of the edge. For decades, CDNs have built large private networking paths and local caching and storage critical for downstream data, such as broadcasting video. Akamai, which leads the CDN market by far, claims that it resides within one network hop for more than 90% of internet users globally. But the opportunity is not just about data: Akamai and other CDN providers have recently diversified into edge computing and are offering “edge cloud platforms” that enable enterprises to run virtualized workloads over their networks. Initial offerings include security services and serverless computing (which lets customers run code without having to provision or manage servers); more are expected.

Networking giants are also re-positioning. Ericsson, Nokia and Huawei have each formed “edge cloud networks” that link points of presence (PoPs) inside telecom “central offices” and colos with last-mile networks, including from wired and wireless telecoms partners. They now — and will likely increasingly — offer edge computing services of their own.

Micro data centers, software-defined

New capacity at the local edge will also be needed, including in small micro data centers that will hand off a portion of edge data (and state) to core/cloud platforms running in larger facilities (see **Micro data centers: An explosion in demand, in slow motion**). Suppliers are lining up. They range from firms exploring multi-tenant micro data center models, such as Vapor IO and startup EdgeInfra, to edge property development investment companies such as SmartEdge DC, which is focusing on heat-reuse models with other buildings.

Edge micro data centers will be managed remotely by software and artificial intelligence (AI) to orchestrate edge workloads in a programmable way. Colos, clouds, carriers and specialists are focusing on new capabilities. They include startups Macrometa, which maintains stateful instances of globally distributed applications; and Rafay Systems, which monitors the performance of distributed applications and, if issues are detected, will automatically create local instances elsewhere.

Demand for edge micro data centers will likely be tempered by the growing power and sophistication of IT appliances at the edge. Chip companies are developing new types of multi-core processors for fast, intelligent IoT gateways and appliances that will likely process the vast bulk of the new edge data created. Edge chips will also be used (eventually) to manage the routing of 5G network traffic across massive multi-antenna arrays.

Big cloud at the edge

Big cloud computing providers will also play a role. Hyperscale data centers in the core will continue to grow, driven by enterprise applications, archiving and other workloads, and so will cloud-to-cloud traffic. All are jockeying for early edge positions with proprietary edge cloud platforms. AWS has been particularly aggressive. It continues to add new regions to its CloudFront CDN, and about two years ago launched AWS IoT Greengrass software to run local compute, storage and messaging on connected devices, such as cameras and edge gateway appliances.

The big public cloud providers are also offering servers and/or software to run cloud workloads in enterprise on-premise locations, including colos — a clear sign that high networking costs and high data transport and bandwidth requirements are best addressed by local processing and storage capacity. AWS Outposts, Microsoft Azure Stack and Google Cloud Services Platform are all designed to run in privately owned physical environments (although inside the cloud provider's firewall). They will be well-suited to work in highly distributed edge environments.

> Adoption of the internet of things, 5G and other edge applications is set to have a dramatic impact, with more applications processing, more switching, and more data never reaching the hyperscale core.

TREND THREE

Data center energy use goes up and up

Energy use by data centers and IT will continue to rise, putting pressure on energy infrastructure and raising questions about carbon emissions. The drivers for more energy use are simply too great to be offset by efficiency gains.

Demand for digital services has seen sustained, exceptional growth over the past few years — and with it, the energy consumption of the underlying infrastructure has risen steadily as well. This has given rise to concerns about the ability of the energy industry to effectively supply data centers in some geographies — and the continuing worries about the sector's growing carbon footprint.

Although there is a shortage of reliable, comprehensive data about the industry's use of energy, it is likely that some models have underestimated energy data and carbon emissions and that the issue will become more critical in the years ahead.

There are some standout examples of IT energy use. Bitcoin mining, for example, is reliably estimated to have consumed over 73 terawatt-hour (TWh) of energy in 2019. This equates to the electricity use of 6.8 million average US households, or 20 million UK households. This is one cryptocurrency — of over 1,500 — and just one application area of blockchains.

Social media provides another example of uncontrolled energy use. Research by Uptime Intelligence shows that every time an image is posted on Instagram by the Portuguese soccer star Cristiano Ronaldo (who at the time of writing had the most followers on the platform), his more than 188 million followers consume over 24 megawatt-hours (MWh) of energy to view it.

Media streaming, which represents the biggest proportion of global traffic and which is rising steadily and globally, has become the energy guzzler of the internet. According to our analysis, streaming a 2.5 hour high definition (HD) movie consumes 1 kilowatt-hour (kWh) of energy. But for 4K (Ultra HD) streaming — expected to become more mainstream in 2020 — this will be closer to 3 kWh, a three-fold increase.

Data from the most developed countries shows what can be expected elsewhere. In the UK, which has more than 94% internet penetration, annual household broadband data consumption increased from 17 gigabyte (GB) in 2011 to 132 GB in 2016, according to official Ofcom data — a sustained 50% increase year-on-year for five years. (The growth figure is much higher in other parts of the world such as Asia and Africa.) Internet penetration, standing at 58% globally in 2019, is expected to increase by 10% in 2020.

This increase in demand is a big driver — although not the only one — for more infrastructure and more energy consumption in cloud, colocation and some enterprise data centers. But a new factor has yet to kick in: 5G.

While it will take a few years for 5G to further mature and become widespread, it is widely expected that the rollout of 5G from 2020 will substantially accelerate the data growth trends, with many new types of digital services in domains such as smart cities, IoT and transportation, among many others. The increased bandwidth compared with 4G will lead to increased demand for higher resolution content and richer media formats (e.g., virtual reality) as soon as late 2020 and rising more steeply, along with energy consumption, after that.

The role of blockchain (of which Bitcoin is just an example) and its impact on energy consumption is still to be fully determined, but if the takeup is on a large scale, it can only be an upward force. Most analysts in this area have predicted a dramatic rise in blockchain adoption beyond cryptocurrency in 2020, helped by new offerings such as the AWS blockchain service. Not all blockchain models are the same, but it inherently means a decentralized architecture, which requires extensive infrastructure to accommodate the replication of data. This consumes more energy than traditional centralized architectures.

Bitcoin is an example of a blockchain that uses Proof of Work as a consensus mechanism — and such models are extremely energy-intensive, requiring multiple parties to solve complex mathematical problems. While alternatives to this model (e.g., Proof of Stake) are likely to gain widespread commercial adoption, the uptake to date has been slow.

Energy consumption and global IT

Several reports have been published in recent years on IT energy consumption and its predicted growth rates. An International Energy Agency (IEA) report published in 2019 noted that workloads and internet traffic will double, but it also forecast that data center energy demand will remain flat to 2021, due to efficiency trends. It cited various references for the basic research.

SELECTED REPORTS ON DATA CENTER INDUSTRY ENERGY USE FROM THE PAST FIVE YEARS

Date	Report	Predictions and calculations*
2014	European Union FP7 Pan-European Data Centre Academy (PEDCA) project	The energy consumption of the European data center industry was approximately 104 TWh in 2014.
2015	On global electricity usage of communication technology: Trends to 2030 (Challenges 2015)	Data center electricity consumption will likely increase by about 15-fold by 2030, representing 8% of projected global electricity demand.
2016	United States data center energy usage report (Lawrence Berkeley National Laboratory)	US data centers are projected to consume around 73 TWh in 2020, a 4% increase compared with 2014.
2017	European Union (EU) H2020 EU Resource Efficiency Coordination Action (EURECA) project	The energy consumption of the European data center industry was around 130 TWh in 2017, a 25% increase compared with 2014.
2019	Powering the cloud: How China's internet industry can shift to renewable energy (Greenpeace)	Electricity consumption of China's data center industry was 160.89 TWh in 2018 and expected to reach 266.79 TWh by 2023.
2019	Tracking clean energy progress: Data centres and data transmission networks (International Energy Agency)	Electricity consumption of the global data center industry was 197.8 TWh in 2018 and is expected to dip slightly to 190.1 TWh by 2021.

* TWh - terawatt-hour
Source: Uptime Institute Intelligence, October 2019

Uptime Institute® | INTELLIGENCE

But Uptime Institute Intelligence is wary of this prediction and intends to collaborate with various parties in 2020 to research this further. There are very strong factors driving up IT energy consumption, and some of the existing data on IT energy use contradicts the IEA figures. The IEA report, for example, stated that global data center energy consumption was 197.8 TWh in 2018 and is expected to drop slightly by 2021. However, research by the European Union's (EU's) EURECA (EU Resource Efficiency Coordination Action) Project found that European data centers consumed 130 TWh in 2017, whereas Greenpeace put energy consumption by the Chinese data center industry at 160 TWh in 2018. This suggests an annual total for China and Europe alone in the neighborhood of 290 TWh, far higher than the IEA global figures.

It is true that the explosive increase in IT demand will not translate directly into the same rate of growth for infrastructure energy consumption (due to increased IT energy efficiency). However, given the exponential rate of growth, it is likely that demand will substantially outpace the gains from efficiency practices over the next five years.

In US data centers, the law of diminishing returns may begin to limit the impact of energy savings. For example, at the data center level, best practices such as hot/cold aisle containment, installation of blanking plates and raising set point temperature have already been widely deployed; this can be seen in the substantial drop in power usage effectiveness (PUE) between 2011 and 2014. However, since 2014, PUE has not dropped much, and in 2019, we noticed a slight increase in the average annual PUE reported by respondents to our global data center survey. Similarly, with IT hardware, Moore's law has slowed down, and newer servers are not maintaining the same efficiency improvements seen in the past.

Uptime Institute expects the strong growth in the IT sector to be sustained over the next five years, given the well-understood demand patterns and the existing technologies coming into large-scale adoption. Our preliminary research suggests that IT energy consumption will rise steadily too, by as much as 10% in 2020, but further research will be conducted to develop and validate these forecasts.

TREND FOUR

Capital inflow boosts the data center market

Data centers are no longer a niche or exotic investment among mainstream institutional buyers, which are swarming to the sector. There is now a buyer for almost every type of data center.

The number of data center acquisitions in 2019 will likely end up being a record high. Among the most active buyers to date are data center companies, such as cloud, colocation and wholesale providers, and data center real estate investment trusts (REITs), as well as private equity firms. In 2019, however, these buyers were increasingly bidding against buyers that, until recently, used to be rare in the sector: infrastructure funds and sovereign wealth funds.

How might this new mix of capital sources change the broader data center sector?

Traditional buyers will remain active

Data center companies — meaning cloud, colocation, wholesale and hosting companies — will continue to buy and sell; Digital Realty Trust, for example, recently purchased Interxion for a whopping $8.4 billion, cementing its place as the largest acquirer in the sector (having invested $22.75 billion since 2010).

These traditional buyers typically acquire data centers for strategic reasons, which means they tend to hold onto acquired assets. Most data center companies are buying to expand their geographic footprint. For colos, expansion is not just to reach more customers but also to be more appealing to large cloud providers. Big cloud customers are prized tenants because they attract additional customers, including enterprises, carriers and service providers. They also tend to be long-term tenants; even if a cloud provider builds its own data center in a region, it will usually continue to lease space, provided costs are low, to avoid migration costs and challenges.

These data center companies will continue to be active acquirers over the next several years, in line with market demand. Facilities with rich fiber connectivity and/or in locations with high demand but limited space will be targets, including in Amsterdam, Frankfurt, Paris, northern Virginia and Singapore. Many will seek to expand in Asia, although to date there has been limited opportunity. (Successful Asia-based data centers are mostly owned by conglomerates, usually with government ties, that seem reluctant to sell.)

But these same data center companies won't only be buying. They see an opportunity to sell off nonstrategic, third-tier assets or those in need of expensive refurbishment to the new financial buyers that take a less strategic view.

For many years, data center companies have also competed with private equity firms for deals. Private equity companies have tended to snap up data centers with a view toward selling them relatively quickly and at a profit. This has driven up the number of acquisition deals in the sector, and in some markets has driven up valuations.

More long-term capital sources move in

So what is changing? More recently, alternative investment firms such as traditional infrastructure investors and sovereign wealth funds have been acquiring data centers. Traditional infrastructure investors, which historically have focused on assets ranging from utility pipelines to transportation projects, have been the most active among new buyers.

The newcomers are similar in that they tend to have longer return periods (that is, longer investment timelines) and lower return thresholds than private equity investors. Many are buying data centers that can be leased, including wholesale for a single large cloud customer and colocation for multiple tenants.

Traditional infrastructure investors, in particular, will likely continue to take over data centers, which they now include as part of their definition of infrastructure. This means they view them as long-term assets that are needed regardless of macroeconomic changes and that provide a steady return. These investors include infrastructure funds, traditional (that is, non-data center) REITS and real estate investors.

In addition to being attracted to the sector's high yields and long-term demand outlook, new investors are also simply responding to demand from enterprises looking to sell their data centers or switch from owning to leasing. As discussed in **Pay-as-you-go model spreads to critical components**, the appetite of enterprises for public cloud, colocation and third-party IT services continues to grow.

An influx of buyers is matching the needs of the sellers. As enterprises outsource more workloads, they need less capacity in their own data centers. Many are (or are in the process of) consolidating their owned footprints by closing smaller and/or regional data centers and moving mission-critical IT into centralized, often larger, facilities. Frequently, the data centers they're exiting are in prime locations, such as cities, where demand for lease space is high, or secondary regional markets that are underserved by colos. All types of investors are buying these enterprise data centers and converting them into multi-tenant operations.

Also common are sales with leaseback provisions ("leasebacks"), whereby an enterprise sells its data center and then leases it back from the new owner, either partially or wholly. Leaseback is attractive to enterprises that want to maintain operational control over their IT but shed the long-term commitment and risk of ownership. Leasebacks also give them more flexibility because they can, for example, only lease a portion of the data center. The new owner benefits because the previous owner becomes an anchor tenant.

Small or regional colo providers are also seeking to sell their data centers, realizing that they lack the economies of scale to effectively compete with multi-national competitors.

More joint ventures

Another trend has been an increase in joint ventures (JVs), particularly by data center REITs, with these new types of investors. For example, in June 2019, the largest data center REIT Equinix formed its first JV to support its new hyperscale data centers (which it will lease to just one or two of the largest cloud providers) with the Singaporean sovereign wealth fund GIC Private Limited, which receives an 80% equity stake. The table below highlights some of the non-traditional capital sources that acquired data centers in 2019.

YOUR DATA CENTER IS OWNED BY ... WHOM?

Examples of data center mergers, acquisitions and joint ventures by new types of investors in 2019, to date

Target	Investor type and name	Details*
AT&T	Infrastructure investor: Brookfield Infrastructure Partners	Closed Jan '19; sale of 31 data centers for $1.1B and formation of Evoque Data Center Solutions. (US)
QTS Realty Trust	Infrastructure investor: Alinda Capital Partners	Announced Feb '19; $53M JV for a hyperscale build. (US)
NTT	Infrastructure investor: GI Partners	Announced Jan '19; sale leaseback of two data centers. (US)
Colo Atl	REIT and wireless firm: American Tower	Announced April '19; data center acquisition. (US)
Digital Realty Trust	Infrastructure investor: Brookfield Infrastructure	Closed April '19; $1.8B JV to buy colo Ascenty. (Brazil)
JM Family Enterprises	Real estate investor: The Woodbery Group	Announced May '19; $10.75M sale of a 3 MW colo. (US)
NLDC	Infrastructure investor: DWS Group GmbH & Co.	Announced May '19; sale of six colo data centers. (Netherlands)
Equinix's xScale	Sovereign wealth fund: GIC Private Limited	Announced June '19; data center JV (80% GIC). (Paris, London)
Digital Bridge	Real estate investor: Colony Capital	Announced July '19; $325M sale, following a JV in 2018. (US)
The Datacenter Group	Infrastructure investor: DWS Group	Announced July '19; merger of 10 colo data centers (65 MW total). (Netherlands)
Telefónica	Infrastructure investor: Asterion Industrial	Closed August '19; $616M (est.) sale of 11 data centers (29 MW total). (US, Latin America and Spain)
Digital Realty Trust	Infrastructure investors: Mapletree Investments, Mapletree Industrial Trust	Announced Sept '19; $1.4B sale of 10 data centers and JV to build three hyperscales. (US)
Netrality Data Centers	Infrastructure investor: Macquarie Infrastructure	Closed August '19; sale including six carrier hotels. (US)
Cologix	Sovereign wealth fund: Mubadala Investment	Announced Sept '19; $500M data center investment from Emirate of Abu Dhabi. (North America)
Expedient	Infrastructure investor: AMP Capital Investors	Announced Oct '19; $500M+ (est.) sale including 11 data centers. (US)

* JV - Joint venture; MW - megawatt; M - million; B - billion
Source: Uptime Institute Intelligence, October 2019

Uptime Institute | INTELLIGENCE

Another data center REIT, QTS Realty, formed a $53 million JV with infrastructure investment fund Alinda Capital Partners in February 2019 to fund a hyperscale wholesale data center that QTS is building in Virginia. In April 2019, data center REIT Digital Realty Trust closed a $1.8 billion JV with infrastructure investor Brookfield Infrastructure, backing Brazil's leading data center company. (Digital Realty owns 51% and Brookfield Infrastructure, 49% of the JV entity.)

We expect more JVs will form, including with more infrastructure funds, to back more very large leased data centers for large cloud providers, which are struggling to expand fast and seek more leasing options (particularly build-to-suit facilities). At the same time, more enterprise data centers will be sold — increasingly to investors with long-term horizons — and converted into multi-tenant facilities.

As more of this type of long-term-horizon money enters the sector, the portion of facilities that are owned by short-term-horizon private equity investors will be diluted. However, even with owners that intend to invest for the long term, data centers are likely to continue to swap hands.

The new types of capital investors in data centers, with their deep pockets and long return timelines, could boost the sector overall. They are likely, for example, to make it easier for any enterprise wishing to give up data center ownership.

TREND FIVE

More data, more automated data centers

Many managers are wary of too much automation and of handing key decisions and operations to machines or outside programmers. But recent advances, including the broad adoption of DCIM and the introduction of AI-driven cloud services, have made this much more likely.

When compared with the workloads they house or to other types of mission-critical facilities, data centers have low levels of automation. Most data centers today have a narrow (but critical) scope of automation — for emergency power scenarios — yet are also now collecting data that would enable them to automate much more.

While talk of automation in the data center often focuses on lights-out operations and even robotics, the most effective (and, in the years ahead, the most likely) technologies to win widespread adoption will involve linking different software systems and controls together, using a wealth of local and aggregated data.

Basic DCIM lays the foundation

Uptime Institute's data center management software maturity model shows the evolution of the use of data center infrastructure management (DCIM) software toward greater automation. DCIM, which collects and analyzes information about a data center's assets and operational status, is the foundation for automation.

Based on the findings of our past two annual surveys, it is clear that DCIM has become a mainstream technology used in most data centers, in spite of its stuttering history. Today, most data centers fall into Level 2 (Reactive) or Level 3 (Proactive) of our maturity model, with limited levels of integration and predictive analysis. However, the recent development of more AI-driven software and cloud services, such as data center management as a service (DMaaS), means more data centers are likely to reach Level 4 and, over time, Level 5. (For more on DMaaS and AI, including use cases, see our report **Very smart data centers: How artificial intelligence will power operational decisions**.)

The best-run data centers deploy both of DCIM's two core features: unified power and environmental monitoring, and asset change and configuration management (which, when combined, comprise a DCIM suite). They also integrate data from DCIM with data from other systems, known as integrated DCIM, or Level 3 of our maturity model. More data center operators are showing an appetite for more advanced capabilities and higher levels of operating efficiency.

Data center management software maturity model

LEVEL	DESCRIPTION	OPERATING EFFICIENCY	SOFTWARE
Level 5: Self-optimizing, autonomic	AI-driven integrated data center and IT management software adjusts data center behavior in real time and makes best use of resources (both data center and IT) according to goals, rules and service requirements throughout its lifecycle.	HIGH	AI-driven, integrated DCIM with automation
Level 4: Optimizing	Data from physical and virtual IT and data center subsystems integrated; data center and IT models used for prediction, service management and multiple views, optimizing in near real time. AI is applied to DCIM-based data lakes for advanced analytics.	MEDIUM	AI-driven, integrated DCIM
Level 3: Proactive	Physical data center equipment characteristics, location and operational status is tracked. Energy and environmental data is used to reduce risks and waste.	MEDIUM	Integrated DCIM
Level 2: Reactive	Software installed to monitor environmentals and equipment power use. Able to adjust basic controls (e.g., cooling) to demand.	LOW	DCIM monitoring
Level 1: Basic	No integration of infrastructure data. Basic monitoring supplied with equipment. Relies on building management system data. Simple alarming, error messaging.	LOW	Ad hoc

UptimeInstitute® | INTELLIGENCE

Integrating DCIM for automation

Once a DCIM suite is up and running, the next phase is typically to integrate it with various other software — and a host of new automated processes become possible. Some of these are shown in the table below. These can involve integration with IT systems; virtual machine and cloud management systems; and energy and business management systems, including customer relationship management, accounting and financial planning.

AUTOMATED PROCESSES, POWERED BY DCIM, ARE BECOMING MORE POPULAR

- Automatically adjusting or moving power loads or workloads according to policy (i.e., availability, maintenance schedules or energy costs/reliability).
- Automated cloud-based resiliency/disaster recovery, which is being driven more by AI.
- Identifying and powering off underused/comatose servers automatically.
- Automated services (provisioning capacity, colocation customer onboarding, audits, etc.)
- Dynamically interacting or exchanging data or power with a power source (such as a utility, renewable energy farm or microgrid), including integration, switching and/or normalizing voltage, currents, etc.

Data center cooling optimization

A trend we often see, and that bears out in our research, is that once a data center team adopts some type of DCIM, they will add more.

Other types of DCIM software, sold separately from the core tools, enable control. The most popular is data center cooling optimization. It uses machine-learning algorithms to learn relationships between variables such as rack temperature; cooling settings, capacity and redundancy; power use; and risk of failure. The software automatically controls cooling units' variable frequency drives (VFDs), by turning units on and off, and by adjusting VFDs and temperature setpoints up or down. While Google, for example, achieved a lot of publicity for its use of AI to automate changes in cooling, it is far from alone: many other organizations have also adopted this technology over the past decade.

Dynamic cooling optimization software is mostly popular among colocation providers. But new AI-driven forecasting capabilities — such as recommended actions and "what if" scenario planning to address challenges around maintenance, reliability and unknown risks; capacity management; and energy use — are likely to interest more data centers.

Operational process automation

Another type of automation DCIM is data center operational management (DCOM) software. This lesser-known cousin of DCIM digitizes and automates operational and management processes, with step-by-step procedures that can be pre-scripted and supported by documents (audit trail), pictures and video. Processes covered include asset management, including maintenance scheduling and tracking; management of and access to documentation (contract, history); staff qualifications; alerting and escalation procedures; interdependencies with other equipment; change management workflow; and root-cause analysis for incidents. Early adopters of this software (there is only a handful of suppliers) say it improves their operations management and can reduce the likelihood of incidents and outages.

Banks and financial services companies have been early adopters of DCOM but it's likely more colos will adopt it and share key data with customers, as a way to differentiate. Many already provide their customers with DCIM-driven online dashboards. Adding DCOM would, for example, enable customers to track the steps of a remote-hands service (from the colo or partner). A few colos are already using DCOM in this way, because when customers have visibility into a remote-hands service, they are more likely to use it.

Micro data centers

Another trend that will lead to more automation is the deployment of more micro data centers. While this will be a long-term trend (see **Micro data centers: An explosion in demand, in slow motion**), suppliers are investing heavily in automation. This is because it's clear that these smaller data center will not be staffed and must be monitored and managed remotely, using software to automate maintenance and break/fix processes. More management software is being embedded in the equipment and many functions are or will be delivered as a cloud service.

Most data centers already collect adequate data to support greater automation. As more managers seek to operate with higher utilization, lower operational costs, and fewer outages and human errors, the case for more automation will become increasingly compelling.

TREND SIX

Data centers without generators: More pilots, more deployments

For most data center owners and operators, diesel and gas generators are a dirty and expensive necessity. But work is gathering pace to drastically reduce the industry's reliance on this 19th century technology.

Diesel (or gas) generators are a necessity for almost all data centers in use today. While most data center operators and owners would like to use a cleaner, more modern technology, none other than generators so effectively offer low operating costs, high power density, reliability, local control and, as long as fuel can be delivered, open-ended continuous power.

Is this about to change? Not wholly, not immediately and not dramatically — but yes, significantly. Even where not eliminated entirely, it is likely that more projects from 2020 onward will involve less generator cover, driven by the following four areas of activity.

Fuel cells and on-site continuous renewables

The opportunity for replacing generators with fuel cells has been intensively explored (and to a lesser extent, tried) for a decade. At least three suppliers — Bloom Energy (US), Doosan (South Korea) and SOLIDPower (Germany) — have some data center installations. Fuel cells are arguably the only technology, after generators, that can provide reliable, on-site, continuous power at scale today.

However, some — including the city of Santa Clara in California — maintain that fuel cells, like generators, are not clean and green. Others say that using grid-supplied or local storage of gas introduces risks to availability and safety.

These objections are much debated and possibly overcome, given the relatively low carbon emissions of fuel cells, the reliability of gas and the fact that very few safety issues ever occur. But price remains an issue: fuel cells do cost more than generators on a kilowatt-hour per dollar ($) basis and have mostly proven economic only when supported by grants. They also require a continuous, steady load (depending on the fuel cell architecture), which causes design and cost complications.

Even so, fuel cells are being deployed: a planned data center campus in Connecticut (owner/operator currently confidential) will have 20 MW of Doosan fuel cells, Equinix is committing to more installations, and Uptime Institute is hearing of new plans elsewhere. The overriding reason is not cost or availability, but rather the ability to achieve a dramatic reduction in carbon dioxide and other emissions and to build architectures in which standby power equipment is not sitting idle.

Edge data centers

Many smaller data centers, perhaps below 500 kilowatt (kW), are expected to be deployed in the decade ahead. Such data centers may more easily duplicate their loads and data to similar data centers nearby, may participate in distributed recovery systems, and may, in any case, cause fewer problems if they suffer an outage.

But above all, these data centers can deploy batteries (or small fuel calls) to achieve a sufficient ride-through time while the network redeploys traffic and workloads. For example, a small shipping container-sized 500 kWh lithium-ion (Li-ion) battery could provide all uninterruptible power supply (UPS) functions, feed power back to the grid and provide several hours of power to a small data center (say, 250 kW) in the event of a grid outage. As the technology improves and prices drop, such deployments will become commonplace.

Cloud-based resiliency

The holy grail for the hyperscale operators, and even smaller clusters of data centers, is to use availability zones, traffic switching, replication, load management and management software to rapidly reconfigure if a data center loses power.

Such architectures are proving effective to a point, but they are expensive, complex and far from fail-safe. All of the major operators continue to build data centers with concurrent maintainability and on-site power at the data center level.

But as software improves and processing/memory falls in price, will this change? Based on the state of the art in 2019 and the plans for new builds, the answer is categorically “not yet.” Even so, expect pioneers to keep trying, especially for noncritical loads, where limited degradation of performance is acceptable: in 2019, at least one major operator conducted tests to determine its resiliency using these technologies. The likely goal would not be to eliminate generators altogether, but rather to reduce the portion of the workload that would need generator cover.

Lithium-ion batteries and smart energy

From 2010 to 2018, the cost of Li-ion batteries (in $ per kWh) fell 85%, according to BloombergNEF (New Energy Finance). Most analysts expect prices to continue to fall steadily for the next five years, with large-scale manufacturing being the major reason. This is creating an opportunity to introduce a new form of energy storage in new ways — including the replacement of some generators.

It is early days, but major operators, manufacturers and startups alike are all looking at how they can use Li-ion storage, combined with multiple forms of energy generation, to reduce their reliance on generators. Perhaps this should not be seen as the direct replacement of generators with Li-ion storage, since this is not likely to be economic for some time, but rather the use of Li-ion storage not just as a standard UPS, but more creatively and more actively. Trials and pilots in this area are likely to be initiated or publicized in 2020 or soon after.

(Alternative technologies that could compete with lithium-ion batteries in the data center include sodium-ion batteries based on Prussian blue electrodes.)

Generators: Still essential for most, but not for all

PROS	CONS
• High power density	• High carbon dioxide emissions
• Reliable, on-site power source	• High particulate emissions
• Proven at scale, over time	• High capital costs
• Available technology, skills common	• Subject to regulation, operating limits
• Indefinite continuous on-site power source	• Assets unused except during emergency
• Low/medium operating costs	• Disciplined testing, maintenance needed
• Can be easily synchronized in series, parallel	• Large, on-site fuel tanks needed
	• Requires significant space
	• Not dynamic or easily redeployed
	• Power is not available instantaneously, requiring separate UPS*

* UPS - Uninterruptible power supply

UptimeInstitute | INTELLIGENCE

Most big data centers cannot realistically contemplate designing or operating data centers to run economically and reliably without generators — yet. But there is a strong drive to do so. Technological alternatives are improving, and the number of good use cases is proliferating. The next 24 months are likely to yield more pilots and deployments.

TREND SEVEN

Pay-as-you-go model spreads to critical components

As enterprises continue to move from a focus on capital expenditures to operating expenditures, more data center components will also be consumed on a pay-as-you-go, "as a service" basis.

The trend toward everything "as a service" (XaaS) is now mainstream in IT, ranging from cloud (infrastructure-aaS) and software-aaS (SaaS) to newer offerings, such as bare metal-aaS, container-aaS, and AI-aaS. At the IT level, service providers are winning over more clients to the service-based approach by reducing capital expenditures (capex) in favor of operational expenditures (opex), by offering better products, and by investing heavily to improve security and compliance. More organizations are now willing to trust them.

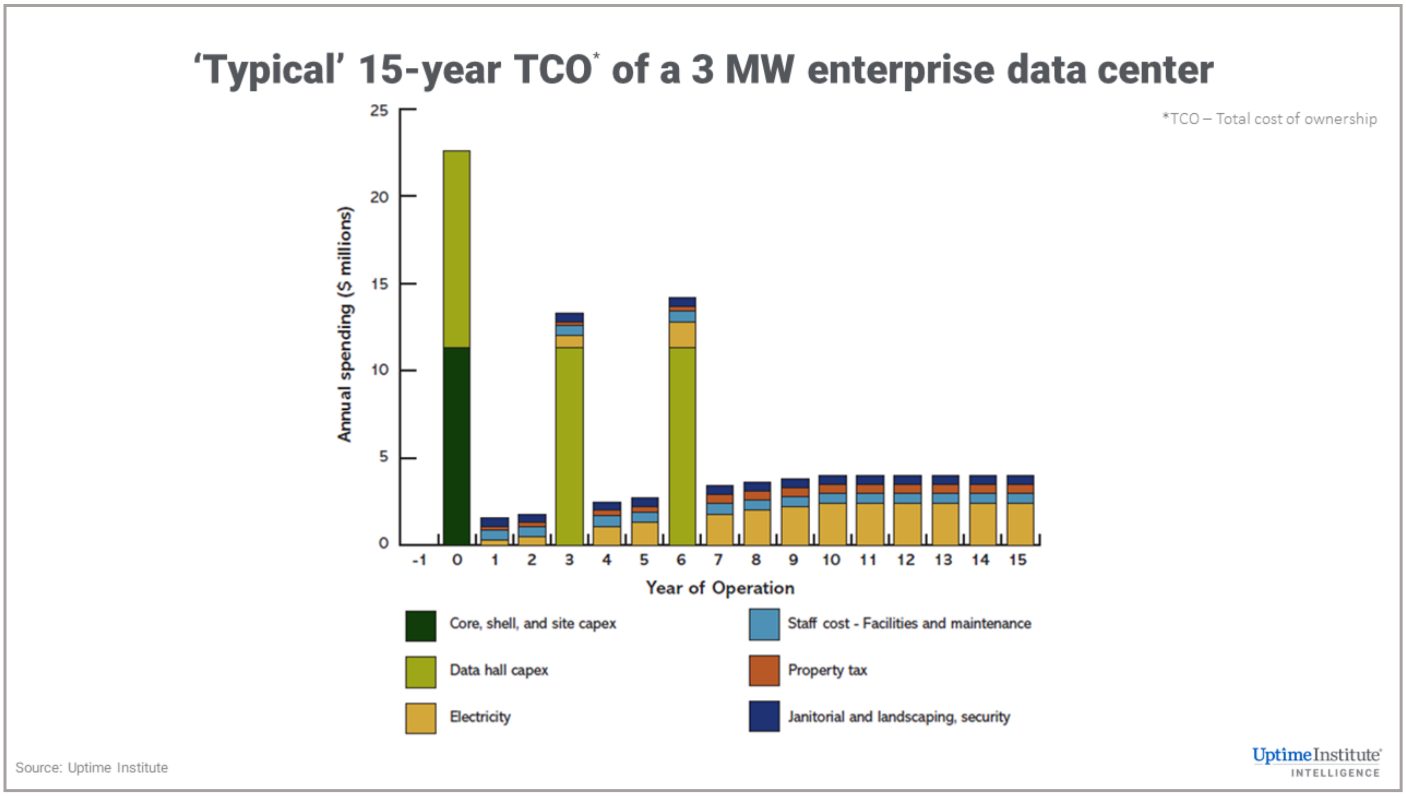
But this change is not confined to the IT: a similar trend is underway in data centers.

Why buy and not build?

While the cost to build new data centers is generally falling, driven partly by the availability of more prefabricated components, enterprise operators have been increasingly competing against lower-cost options to host their IT — notably colocation, cloud and SaaS.

Cost is rarely the biggest motivation for moving to cloud, but it is a factor. Large cloud providers continue to build and operate data centers at scale and enjoy the proportional cost savings as well the fruits of intense value engineering. They also spread costs among customers and tend to have much higher utilization rates compared with other data centers. And, of course, they invest in innovative, leading-edge IT tools that can be rolled out almost instantly. This all adds up to ever-improving IT and infrastructure services from cloud providers that are cheaper (and often better) than using or developing equivalent services based in a smaller-scale enterprise data center.

Many organizations have now come to view data center ownership as a big capital risk — one that only some want to take. Even when it's cheaper to deliver IT from their own "on-premises" data center, the risks of data center early obsolescence, under-utilization, technical noncompliance or unexpected technological or local problems are all factors. And, of course, most businesses want to avoid a big capital outlay: Our research shows that, in 2017, the total cost of ownership of an "average" concurrently maintainable 3 megawatt (MW) enterprise data center amortized over 15 years was about $90 million, and that roughly half of the cost is invested in three installments over the first six years, assuming a typical phased build and bricks-and-mortar construction.



This represents a significant amount of risk. To be economically viable, the enterprise must typically operate a facility at a high level of utilization — yet forecasting future data center capacity remains enterprises' top challenge, according to our research.

Demand for enterprise data centers remains sizable, in spite of the alternatives. Many enterprises with smaller data centers are closing them and consolidating into premium, often larger, centralized data centers and outsourcing as much else as possible.

The appeal of the cloud will continue to convince executives and drive strategy. Increasingly, public cloud is an alternative way to deliver workloads faster and cheaper without having to build additional on-premise capacity. Scalability, portability, reduced risk, better tools, high levels of resiliency, infrastructure avoidance and fewer staff requirements are other key drivers for cloud adoption. Innovation and access to leading-edge IT will likely be bigger factors in the future, as will more cloud-first remits from upper management.

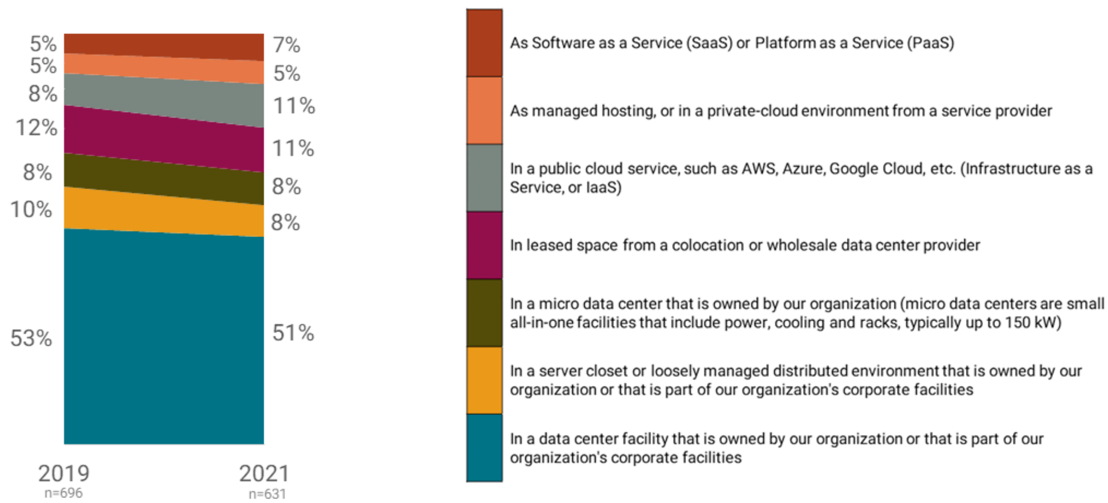
Colocation, including sale leasebacks

Although rarely thought of in this way, colocation is the most widely used "data center-aaS" offering today. Sale with leaseback of the data center by enterprise to colos is also becoming more common, a trend that will continue to build (see **Capital inflow boosts the data center market**).

Colo interconnection services will attract even more businesses. More will likely seek to lease space in the same facility as their cloud or other third-party service provider, enabling lower latency and fewer costs and more security for third-party services, such as storage-aaS and disaster recovery-aaS.

While more enterprise IT is moving to colos and managed services (whether or not it is cloud), enterprise data centers will not disappear. More than 600 IT and data center managers told Uptime Institute that, in 2021, about half of all workloads will still be in enterprise data centers, and only 18% of workloads in public cloud/SaaS.

Corporate IT has many venues



Approximately what percentage of your organization's total IT would you describe as running in the following IT environments today versus in two years? (Your answers for each year must sum to 100%)

Source: Uptime Institute Global Survey of IT and Data Center Managers 2019

UptimeInstitute | INTELLIGENCE

Other "as a service" trends in data centers

Data center monitoring and analysis is another relatively new example of a pay-as-you-go service. Introduced in late 2016, DMaaS is a big data-driven cloud service that provides customized analysis and is paid for on a recurring basis. The move to a pay-as-you-go service has helped unlock the DCIM market, which was struggling for growth because of costs and complexity.

Energy backup and generation is another area to watch. Suppliers have introduced various pay-as-you-go models for their equipment. These include leased fuel cells owned by the supplier (notably Bloom Energy), which charges customers only for the energy produced. By eliminating the client's risk and capital outlay, it can make the supplier's sale easier (although they have to wait to be paid). Some suppliers have ventured in UPS-aaS, but with limited success to date.

More alternatives to ownership are likely for data center electrical assets, such as batteries. Given the high and fast rate of innovation in the technology, leasing large-scale battery installations delivers the capacity and innovation benefits without the risks.

It's also likely that more large data centers will use energy service companies (ESCOs) to produce, manage and deliver energy from renewable microgrids. Demand for green energy, for energy security (that is, energy produced off-grid) and energy-price stability is growing; ESCOs can deliver all this for dedicated customers that sign long-term energy-purchase agreements but don't have the capital required to build or the expertise necessary to run a green microgrid.

WHY PAY-AS-YOU-GO AND "AS A SERVICE" ARE GAINING GROUND	AND WHY IT ISN'T ALWAYS BEST ...
• Reduces risk by taking large capex items of the balance sheet • Enables innovation without large capital expenditure • Reduces lock-in due to depreciation, etc. • Allows buyers to move between services (in theory) • Enables suppliers to spread the costs of development, facilities across clients • Reduces exposure to staff, skills shortage risks • Reduces the length and cost of sale for the supplier, because there is no large capital outlay	• Full privacy and control can best be ensured by ownership • There can be tax advantages to ownership • Economies of scale don't always apply for customized requirements • Development is driven by group needs, or supplier needs, not individual customers • Service quality, transparency and future investment cannot be guaranteed • Compliance issues might arise • Some service providers may be weaker on security than their customers

UptimeInstitute | INTELLIGENCE

Demand for enterprise data centers will continue but alongside the use of more cloud and more colo. More will be consumed "as a service," ranging from data center monitoring to renewable energy from nearby dedicated microgrids.

TREND EIGHT

Micro data centers: An explosion in demand, in slow motion

A wave of new technologies, from 5G to the IoT to AI, means much more computing and much more data will be needed near the point of use. That means many more small data centers will be required. But there will be no sudden mass deployment, no single standout use case, no single design dominating. Demand is likely to grow faster from 2022.

Suppliers in the data center industry are excited. Big vendors such as Schneider, Vertiv and Huawei have been rapidly adding to their product lines and redrawing their financial forecasts; startups — companies such as Vapor IO, EdgeMicro, EdgeInfra and MetroEDGE — are pioneering new designs; and established telco specialists, such as Ericsson, along with telco operators, are working on new technologies and partnerships. Builders and operators of colocation data centers, such as EdgeConneX, Equinix and Compass, are assessing where the opportunity lies.

The opportunity is to supply, build or operate local edge data centers — small micro data centers that are designed to operate near the point of use, supporting applications that are not suited to run in big, remote data centers, even in mid-sized regional colocation data centers. Unlike most larger data centers, micro data centers will mostly be built, configured and tested in a factory and delivered on a truck. Typical sizes will be 50 kW to 400 kW, and there are expected to be a lot of them.

But with the anticipation comes consternation — it is possible to commit too early. Some analysts had predicted that the explosion in edge demand would be in full swing by now, fueled by the growing maturity of the IoT and the 2020 launch schedules for 5G services. Suppliers, however, mostly report only a trickle — not a flood — of orders.

Privately, some suppliers admit they have been caught off guard. There is a deep discussion about the extent of data center capacity needed at the local edge; about just how many applications and services really need local edge processing; and about the type and size of IT equipment needed — maybe a small box on the wall will be enough?

While the technical answers to most of these questions are largely understood, questions remain about the economics, the ownership, and the scale and pace of deployment of new technologies and services. These are critical matters affecting deployment.

Edge demand and 5G

In the past decade, data and processing has shifted to a cloudy core, with hundreds of hyperscale data centers built or planned. This will continue. But a rebalancing is underway (see **The internet tilts toward the edge**), with more processing being done not just at the regional edge, in nearby colocation (and other regional) data centers, but locally, in a micro data center that is tens or hundreds of meters away.

This new small facility may be needed to support services that have a lot of data, such as MRI scanners, augmented reality and real-time streaming; it may be needed to provide very low latency, instantly responsive services for both humans and machines — factory machines are one example, driverless cars another; and it may be needed to quickly crunch AI calculations for immediate, real-time responses. There is also a more mundane application: to provide on-site services, such as in a hospital, factory or retail establishment, should the network fail.

With all these use cases, why is there any doubt about the micro data center opportunity?

First, in terms of demand drivers, no new technology has created so much interest and excitement as 5G. The next generation telecom wireless network standard promises speeds of up to 10 gigabits per second (Gbps) communications, latency of below five millisecond (ms), support for one million devices per square kilometer, and five-nines availability. It will ultimately support a vast array of new always-on, low latency and immersive applications that will require unimaginable amounts of data and compute power — too much to realistically or economically send back to the internet's hyperscale core. Much of this will require low-latency communications and rapid processing of a few milliseconds or less — which, the speed of light dictates, must be within a few kilometers.

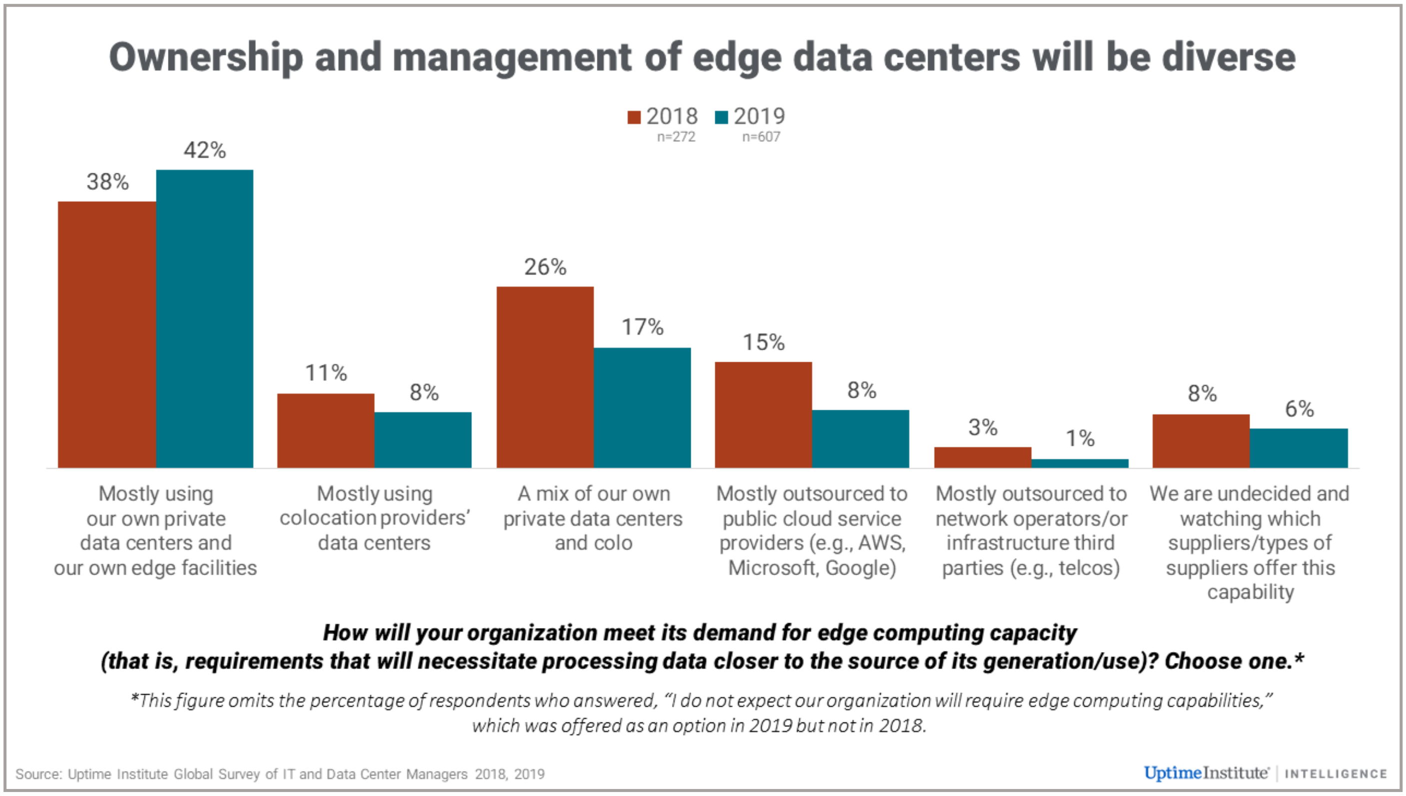
Few doubt that 5G will create (or satisfy) huge demand and play a pivotal role in IoT. But the rollout of 5G, already underway, is not going to be quick, sudden or dramatic. In fact, full rollout may take 15 years. This is because the infrastructure required to support 5G is too expensive, too complex, and involves too many parties to do all at once. Estimates vary, with at least one analyst firm predicting that telecom companies will need to spend $1 trillion upgrading their networks.

A second issue that is creating uncertainty about demand is that many edge applications — whether supported by 5G or some other networking technology (such as WiFi 6) — may not require a local micro data center. For example, high-bandwidth applications may be best served from a CDN at the regional edge, in a colo, or by the colo itself, while many sensors and IoT devices produce very little data and so can be served by small gateway devices. Among 5G's unique properties is the ability to support data-heavy, low-latency services at scale — but this is exactly the kind of service that will mostly be deployed in 2021 or later.

Suppliers and telcos alike, then, are unsure about the number, type and size of data centers at the local edge. Steve Carlini, a Schneider Electric executive, told Uptime Institute that he expects most demand for micro data centers supporting 5G will be in the cities, where mobile edge-computing clusters would likely each need one micro data center. But the number of clusters in each city, far fewer than the number of new masts, would depend on demand, applications and other factors.

A third big issue that will slow demand for micro data centers is economic and organizational. These issues include licensing, location and ownership of sites; support and maintenance; security and resiliency concerns; and management sentiment. Most enterprises expect to own their own edge micro data centers, according to Uptime

Intelligence research, but many others will likely prefer to outsource this altogether, in spite of potentially higher operational costs and a loss of control.



Suppliers are bullish, even if they know demand will grow slowly at first. Among the first-line targets are those simply looking to upgrade server rooms, where the work cannot be turned over to a colo or the cloud; factories with local automation needs; retailers and others that need more resiliency in distributed locations; and telcos, whose small central offices need the security, availability and cost base of small data centers.

This wide range of applications has also led to an explosion of innovation. Expect micro data centers to vary in density, size, shape, cooling types (include liquid), power sources (including Li-ion batteries and fuel cells) and levels of resiliency.

The surge in demand for micro data centers will be real, but it will take time. Many of the economic and technical drivers are not yet mature; 5G, one of the key underlying catalysts, is in its infancy. In the near term, much of the impetus behind the use of micro data centers will lie in their ability to ensure local availability in the event of network or other remote outages.

TREND NINE

Staffing shortages are systemic and worsening

The demand for data center staff continues to exceed supply, and more employers are struggling with talent shortages. The sector's people problem is systemic and long term and is likely, among other things, to spur a wave of investment in education and training.

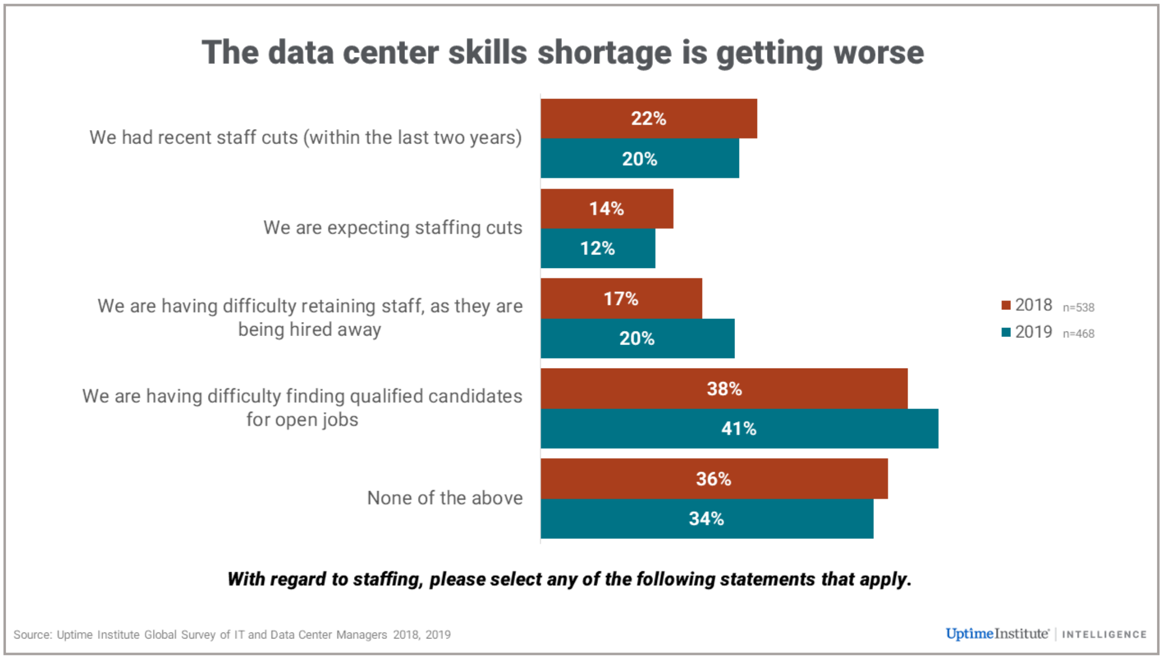
Difficulties recruiting and retaining data center staff has been a growing issue for at least a decade. As demand for data center capacity continues to swell, the problem is intensifying. In our 2019 global survey of IT and data center managers, 61% reported having trouble finding or retaining staff, which is up from 55% a year ago, and 41% are having difficulty finding qualified candidates for open jobs, up slightly from 38%.

Our research shows that the impact of shortages varies by location. Managers of facilities in rural or semi-rural areas can struggle to attract qualified candidates at any level. Those operating in data center hubs compete to recruit and retain qualified people, especially at senior levels. We hear of large hyperscale operators attracting much of the available talent, which some believe is driving up the cost of compensation packages (although, simultaneously, the salaries at some enterprise and colocation data centers have failed to keep pace with general market rates).

Causes and effects

What we hear anecdotally, and what our research bears out, is that several factors are at play, notably:

- **Demand exceeds supply:** The data center industry has been expanding so rapidly, especially in some geographies, that training, recruitment and relocation of staff simply cannot keep up. Those in



the fastest growing sectors (e.g., cloud) are recruiting staff away from those growing more slowly (e.g., enterprises), which typically pay less.

- **Invisibility:** The general public has little or no familiarity with data centers. As such, the pipeline of people coming into the sector is low; very few students and job seekers consider — or are preparing for — a data center career.
- **Lack of diversity:** Gender and other demographic imbalances deter some potential candidates and can create staff retention issues. (Our research shows that women comprise less than 5% of data center facilities teams. In 2019, 45% of survey respondents said that they believe a lack of women posed a threat to the industry, up from 30% a year earlier.)
- **A graying workforce:** Most of our survey respondents (56%) have spent more than 20 years in the sector and we know many senior staff are set to retire within a short timeframe, especially in smaller data center operations. (About one-fifth — 21% — of our 2019 survey respondents have spent 30 years or more in the sector.)
- **Tight labor markets:** In leading data center hub nations such as China, Malaysia, Singapore, the US and the UK, unemployment rates are 4.0% or lower. Ireland's 4.6% unemployment rate is also low.

For the most part, employers have been coping with shortages by adjusting salaries and investing in new recruiting efforts (including at educational institutions) and in more on-the-job-training and cross-training. Some are also rewriting job descriptions to be more inclusive and with fewer "hard" requirements. When these efforts fall short, the most common impacts, according to our research, include:

- Increased overtime.
- Heavier workloads.
- Difficulty scheduling.
- Deferred maintenance.
- Reduced management and operations focus.
- Impaired knowledge transfer (especially at senior levels).
- Greater outsourcing.

When positions remain unfulfilled for an extended period, data center personnel budgets are often cut. This creates a vicious cycle whereby overtime and heavier workloads become part of the "standard" working environment, which can, in turn, drive existing staff to seek employment elsewhere.

Savvy employers are developing staffing plans to help retain their personnel budgets, as well as succession plans at all levels. These plans include greater investment in training, including certification and credentialing programs to support professional growth.

Recognizing that no single company, no matter how large, can solve the staffing challenge alone, several partnerships and member organizations are working together to raise awareness of the data center sector and to encourage new talent. Notable examples include the 7x24 Exchange's declaration of October 29, 2019, as the first annual International Data Center Day and the Infrastructure Masons' scholarship program, which has raised more than $250,000 in recent years.

Greater action needed

Increased awareness, coupled with hiring and retention efforts both at the individual company and sector level, will not produce an immediate reprieve. Many data center employers we interviewed believe that the worse is yet to come: the skills shortage will have greater impacts in the near future.

Technologies such as AI and automation may eventually provide relief. However, the problem is deep and widespread, and the impact of technology will likely be limited; only 29% of our survey respondents believe that AI, for example, will lead to reduced employment relative to workloads within five years.

One clear way to build the talent pipeline is by creating more data center-centric curricula, including at trade schools and higher education institutions. Some organizations are already doing this, including the Institute of Technology in Ireland and, in the US, Southern Methodist University and Marist College. We know of more colleges and universities, including in the US and Europe, that are investigating similar curricula.

Industry groups might help create more education opportunities, similar to the coursework recently launched by the National Consortium for Mission Critical Operations for undergraduates and professionals (available to students at a discounted rate, thanks to funding from the US Department of Labor).

> To effectively address the worsening staffing shortage crisis — to ensure stability and growth in the data center sector — more investment will be needed from industry and educators.

TREND TEN

Climate change spurs data center regulations

Growing awareness of climate change and its accelerating effects is causing regulators to take more notice of data centers. More regulations and reforms will be introduced in 2020 and beyond, addressing energy efficiency, renewable energy and waste reduction.

The global information revolution is viewed by many in the industry as a positive force when it comes to environmental impact, enabling greater efficiencies, reduced costs and decarbonization of many goods and services. But the dramatic growth of the data center industry also has a negative environmental impact in terms of power consumption, particulates from generators and the consumption of other resources (e.g., water) and materials, some of which are rare.

Climate change and carbon emissions is a particular concern, and lawmakers are drawing up new regulations and reforms in an attempt to curb the sector's impact. A number of cities and regions have already introduced new restrictions for data centers. This varies from imposing PUE caps and encouraging design for disassembly to initiating new energy and carbon-reporting requirements and reforming energy markets.

For example, the Shanghai municipality introduced a mandatory requirement for new data centers to have a PUE value of less than 1.3. A cap on PUE for new builds is also in place in Amsterdam, which hosts around 30% of all European data centers.

Earlier this year, Amsterdam also introduced a year-long halt on permits for new data centers while authorities attempt to create a new environmental strategy for the sector — one plan is to mandate that waste heat should be used for district heating. Australia, meanwhile, is currently evaluating the situation to decide on the best course of action. In the US, the city of Santa Clara has imposed a ban on onsite fossil fuel-based generation except for backup generators — a move that is being contested in the courts by fuel cell supplier Bloom Energy (which argues that its natural gas-powered fuel cells are a low-carbon technology).

The "green wave"

This wave of legislation and new planning restrictions is not an entirely new development. There have been and are still many rules, codes of conduct and standards (some mandatory, others not) concerning, for example, carbon reporting, diesel generator emissions and electrical/cooling/design efficiency across the world, and especially in Europe. But there is a clear hardening of attitudes and a greater willingness to act.

We expect that in 2020 we will see more regulations with wider scope and impact compared with previous endeavors. These regulations will aim to increase the energy and material efficiency of data center-related products and services.

One example will be the introduction of the EU eco-design regulation for IT servers and storage devices, effective March 2020. Any products sold in Europe and deemed within scope will have to comply with the new law. The law covers servers with up to four processor sockets and will set out rules that cap idle power consumption. Manufacturers are likely to apply the changes to their products globally.

Under the law, vendors will also be required to publish their equipment thermal performance; optimize designs for easier disassembly and re-use; and to commit to firmware updates to prolong the useful life of devices to reduce waste sent to landfills.

Potential impact on industry

To date, the impact of legislation and other government initiatives has been mixed. Voluntary codes of conduct have arguably trailed the practices of leading operators, and green design initiatives and labeling have been a little more impactful. The most powerful rules have been around safety and local environmental impact. California's Title 24 building rules, for example, prescribe the use of environmentally friendly cooling technologies.

Publishing server operating conditions may help data centers better optimize operating temperature, segregate IT kit based on optimal performance sweet spots, and avoid designing for the highest common denominator (i.e., equipment with the most demanding requirements). An idle cap on servers may help reduce IT energy consumption, given the current industry server utilization levels (between 5% and 25%).

Manufacturing for recycling, paired with the ageing Moore's law, will mean more refurbished/redistributed IT kit will likely be deployed in more data centers. Research suggests that performance gains from refreshing servers are diminishing but can be more readily achieved by upgrading existing IT (e.g., by increasing memory capacity). Hyperscalers, such as Google, have introduced internal initiatives to reuse/redeploy older servers for less-demanding applications and to refurbish/remanufacture where possible to prolong equipment life. Others are likely to follow their example.

Setting limits on PUE for data centers will not always yield the intended results and so has proved controversial. PUE is not an energy efficiency metric, many point out; it does not address IT efficiency, nor does its value fairly reflect all the design/building goals of the infrastructure (e.g., reliability, availability and the need for redundancy). Additionally, fluctuations in IT load can cause misleading changes to the PUE numbers. This leads to confusion, and in some cases, misinformation. But efforts to define and promote more meaningful or accurate measures, such as energy consumption per business-useful transaction (taking into consideration all business requirements and constraints), have not gained much traction.

Utilities and renewables

Not all legislation or policies that will affect data centers is directly aimed at the sector. The energy sector, for example, is under pressure around the world to convert to renewable energy sources, and a rapid and dramatic transition is underway in many countries.

Data centers are coming under pressure to buy renewable power (under the appropriate agreements) and to locate in areas where the energy sources are greener. Greenpeace, which for several years has analyzed the sector, is having an effect. It has produced a number of reports on the environmental impact of data centers, including scorecards on the green performance of leading colocation and cloud providers. The biggest providers now have stated goals to reduce the carbon footprint of their data center portfolios. Facebook, for example, sited a new data center in New Mexico, which recently passed a law mandating that 89% of its energy will be renewable by 2040.

More energy markets in various parts of the world will begin to be reformed, giving large consumers, such as data centers, more options to purchase renewable energy. In 2019, the Taiwanese government amended its rules to allow non-utility companies to directly procure renewable energy. Another example: China is currently in the process of reforming its energy markets to foster the uptake of renewable energy in data centers; its recently launched framework “Guidelines on Strengthening the Construction of Green Data Centers” is designed to encourage the use of direct power purchase and green energy certificates.

RECENT EXAMPLES OF LEGISLATION TO REDUCE ENVIRONMENTAL IMPACT

Amsterdam	• PUE* limits on new data centers • Moratorium on new licenses until environmental impact assessed — city wants heat re-use.
Singapore	• Restrictions on new builds due to land use, energy.
Santa Clara	• On-site generation must use nonfossil fuels.
Shanghai	• New data centers must have PUE* of 1.3 or less.
European Union	• New rules governing server energy use when idle, thermal reporting and recyclability.

* PUE - power usage effectiveness
Source: Uptime Institute Intelligence, October 2019

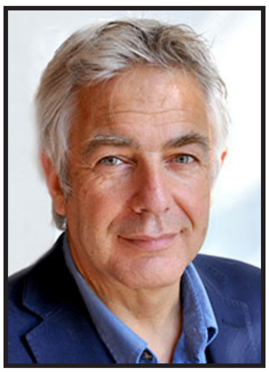
Uptime Institute | INTELLIGENCE

> Although major industry players have started to take bolder actions and reform themselves to reduce their environmental impact, legislators, nongovernmental organizations and the public are pressing for more. Regulatory oversight will increase.

Uptime Institute | INTELLIGENCE

ABOUT THE AUTHORS

Rhonda Ascierto is Uptime Institute's Vice President of Research. 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. Contact: rascierto@uptimeinstitute.com



Andy Lawrence is Uptime Institute's Executive 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. Contact: alawrence@uptimeinstitute.com

Dr. Rabih Bashroush is Uptime Institute's Research Director. An international authority in the field of information and communication technology digital infrastructure and consultant for various European governments (including the European Commission), Dr. Bashroush has led numerous research programs designed to provide a better understanding of digital infrastructure, its impact, and how it can operate more effectively. Contact: rbashroush@uptimeinstitute.com

Kevin Heslin is Chief Editor and Director of Ancillary Projects, Uptime Institute. Kevin served as editor of The Uptime Institute Journal, where he authored and co-authored numerous papers, including case studies involving Uptime Institute Tier-certified facilities. Kevin has had a more than 30-year career examining energy and sustainability issues from the end-user perspective. In that regard, he has worked in academic, nonprofit and standards organizations, in addition to serving as editor of various print publications. Contact: kheslin@uptimeinstitute.com