# Five data center predictions for 2022

## The sector is booming and maturing globally — but management challenges lie ahead

Uptime Institute® | INTELLIGENCE

In this report, Uptime Institute Intelligence looks beyond the obvious trends of 2022 — that the sector is thriving, building out to the edge and recovering and adapting to the pandemic — and identifies some potentially challenging issues. These include sustainability reporting and the difficulty of achieving zero-carbon emissions, the uneven impacts of processor innovations, supply chain problems and the concentration risk of cloud computing.

Authors

**Andy Lawrence,** Executive Director of Research, Uptime Institute

**Rhonda Ascierto,** Vice President of Research, Uptime Institute

**Daniel Bizo,** Research Director, Uptime Institute

**Tomas Rahkonen,** Research Director of Distributed Data Centers, Uptime Institute

25-35 MINUTES TO READ

UII-58 v1.0M published January 06, 2022 last updated January 06, 2022

# Synopsis

The critical digital infrastructure sector continues to expand and attract enviable levels of new investment. The coming year (and beyond) promises opportunities but also challenges that only some data centers will be likely to navigate successfully. This report predicts some of the potential outcomes ahead, including achieving zero-carbon emissions, reducing dependency on cloud, benefiting from processor innovation and ensuring supply-chain resiliency.

© 2022 **UPTIME INSTITUTE** | ALL RIGHTS RESERVED

# Contents



## Illustrations



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

© 2022 UPTIME INSTITUTE | ALL RIGHTS RESERVED

# Introduction

For the digital infrastructure industry, 2021 was an unexpectedly positive year, given the mood over the previous 12 months. When the pandemic struck, in early 2020, the global business outlook was very unsettled, with many data center and IT managers fearing major staffing issues, supply chain problems, budget stresses and a severe recession. It was not clear at this stage that effective vaccines against COVID-19 would be developed or that economies would recover so fast, or even that major political upheaval would be avoided.

Heading into 2022, the mood is — guardedly — more positive. Although the threat of COVID-19 variants persists, we can nevertheless repeat many of the sentiments from our *Five data center trends for 2021*: "The critical digital infrastructure sector continues to expand and to attract enviable levels of new investment. The ongoing build-out of new data centers and networks is largely being driven by cloud, hosted, and other as-a-service workloads…"

Although the threat of COVID-19 variants persists, we can nevertheless repeat many of the sentiments from our trends for 2021.

Overall, the critical infrastructure sector has adapted well to the pandemic. Fears of major outages, equipment shortages, huge budgetary overspends and a high loss of available staff have largely dissipated. These problems have been both more chronic and milder than feared — balanced by a strong uptick in demand as more people and services move online and more investment is made in resiliency and automation.

At the beginning of each calendar year, Uptime Institute highlights a short list of trends — or predictions — for the year ahead. We try to avoid stating the obvious or repeating the forecasts of the previous year. For 2021, we highlighted five trends that have largely been borne out and these are detailed in the **Appendix**. The trends spanned growing demand for more resiliency and accountability; renewed investment in remote monitoring and automation; a belated investment wave in edge data center capacity; a significant increase in sustainability reporting requirements; and, finally, a wave of innovation.

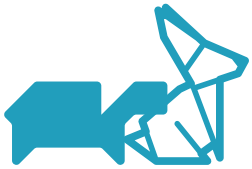
And for 2022? Our five predictions are unusual in that our confidence level in these predicted trends is not as high as usual, and some are also more contentious. However, the purpose of a prediction (or of highlighting a new trend) is not to be exact, but to provide some insights that may help to inform and guide the industry.

© 2022 **UPTIME INSTITUTE** | ALL RIGHTS RESERVED

# 1. Moore's law resumes — but not for all

Sustained advances in chip technology over the past few decades, along with leaps in networking speeds, have helped the data center industry to grow. This has largely been through increased demand for IT services, but also because of vastly improved power efficiencies.

In the past few years, however, the continuous improvements in chip processing speeds and efficiency have stalled as key suppliers, operating at the limits of physics, pushed their manufacturing capability to breaking point. Intel, the global leader in server processors, with over 80% of the market, misjudged some crucial manufacturing technology choices — at a time when there was an insufficient alternative supply. This meant that while raw computing performance per server kept rising, the rate of improvement in IT energy efficiency slowed down — with potentially negative effects on global data center footprint, energy use and costs.



Some data centers will be able to exploit the new technology and make big gains in processing power and energy use — and some will not.

This is likely to change from 2022 onward, with an upswing in performance. However, there will not be a return to past dynamics. Some data centers will be able to exploit the new technology and make big gains in processing power and energy use — and some will not.

A major development is that Intel, whose dominance drives the industry, is set to find its mojo again — but only gradually. In the first months of 2022, the company will start making a new server processor generation, built on an enhanced version of its 10-nanometer technology (branded as Intel 7), to increase performance density. Intel's 2022 server processors will bring improvements in efficiency too — but this will only be the first step in Intel's technology comeback.

By the end of 2022, however, the market will have changed — in a way that some will be able to exploit far better than others. Regardless of how fast Intel recovers from its missteps, it will not regain its former grip on server processors, and data center operators will benefit (directly or indirectly) from a more diverse supply of chips. A resurgent Advanced Micro Devices (AMD), Intel's long-standing rival, has already gained a recent foothold with processors that outperform those of its bigger rival's products across a range of server applications. These processors are delivering much higher performance more efficiently and, during 2022, AMD is expected to introduce increasingly efficient products.

Other chip vendors, such as workload-acceleration specialist NVIDIA, as well as some hyperscale operators that have started building their own cloud server platforms — including Amazon Web Services (AWS) and Chinese cloud giant Alibaba — will be migrating to new chip fabricating technologies with their own latest processor designs. Underpinning these developments are TSMC (Taiwan Semiconductor Manufacturing Company) and Samsung Electronics (headquartered in South Korea), two foundries with cutting-edge manufacturing capabilities.

© 2022 UPTIME INSTITUTE | ALL RIGHTS RESERVED

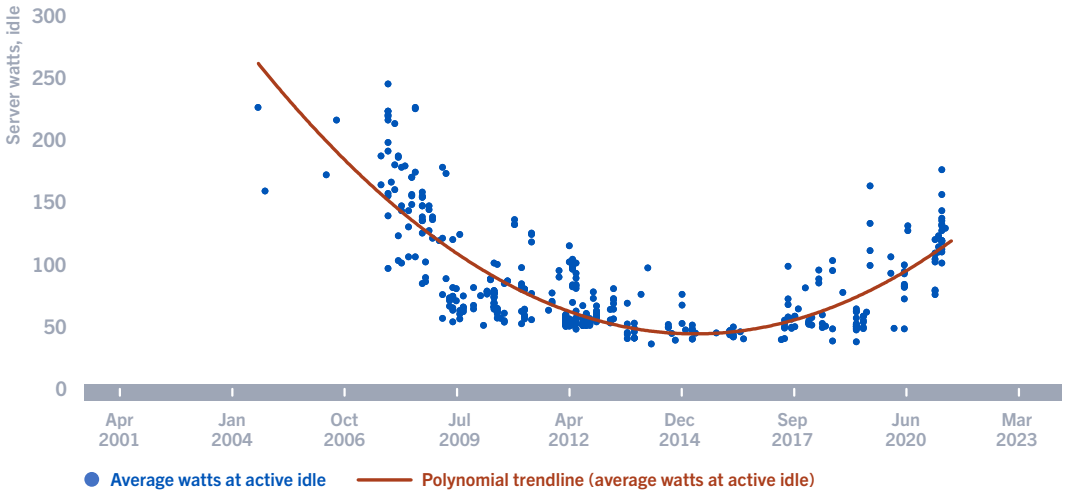
The simultaneous jump in performance and efficiency across multiple computing platforms will stimulate another IT investment cycle in 2022 and 2023, led by powerful IT services providers, such as major cloud, hosting and managed services vendors, hungry for technology upgrades.

These technical leaps, however, will not bring back the carefree ride of the past decades for IT users when servers were continuously improving and sped up applications for all customers. Advances from 2022 onward will benefit at-scale infrastructure operators (and cloud customers indirectly) the most, and the few cutting-edge end users that invest heavily in highly sophisticated IT skills and new software. This is because trends in compute technologies mostly (and increasingly only) favor highly consolidated infrastructure (see below) and software that make use of new features.

So, what does this mean? Infrastructure economics and power efficiency are at their best when operators equip servers with high-performance processors that have many cores and a large memory, and when they are run at high utilization levels. New servers have many cores, and it won't be long before having more than 100 cores will be common — up from a handful 10 years ago. This invites high levels of workload consolidation well beyond best practice up until recently — or, indeed, those of many enterprises today — which will further widen the efficiency gap between cutting-edge and trailing practices.

Another factor is that new servers are no longer becoming more efficient when running light loads (e.g., 10% average utilization or lower). Worse still, idle server power, which has followed a downward trajectory over the past 20 years, has been creeping up (see **Figure 1**). Organizations that do not achieve considerably higher utilization, including through more (preferably dynamic) consolidation, will not see the energy benefits from the latest and future new servers. Cloud services, web technology and high-performance computing users, however, will keep reaping the significant energy efficiency gains seen in the golden past of Moore's law.

**Figure 1** **Historical trending of idle server power**




STANDARD PERFORMANCE EVALUATION CORPORATION, 2021
COMPILED BY UPTIME INSTITUTE

Uptime Institute | INTELLIGENCE

© 2022 UPTIME INSTITUTE | ALL RIGHTS RESERVED

Infrastructure economics will, increasingly, be about much more than utilization through workload consolidation. Newer chips have a growing number of features that can speed up processing but only when supported by the application software. This includes fast low-precision calculations (e.g., for neural networks and data analytics) and ultrawide vector processing; persistent memory pools for large in-memory data sets with several times faster write operations compared with storage systems; or transactional memory for nonblocking data synchronization supporting parallel applications (e.g., database engines, and scientific and technical computing).

There are also the new cloud compute platforms. To benefit from their cost and indirect energy efficiency, IT users need to understand which workloads are appropriate for them and then migrate (or develop) parts of their stack for these systems. To take advantage of the latest workload acceleration techniques in chips requires substantial changes to application code.

Technical organizations with the necessary skills to exploit new chip technology will thrive in infrastructure performance and efficiency in 2022 and beyond. Those without will struggle to improve.

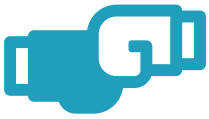
**Advances in chip technology will drive major gains in energy efficiency for data center operators of at-scale infrastructure. Those unable to exploit the improvements or achieve higher utilization through consolidating workloads will struggle to remain competitive.**

© 2022 **UPTIME INSTITUTE** | ALL RIGHTS RESERVED

# 2. Industry consensus on sustainability looks fragile

Pressed by a sense of urgency among scientists and the wider public, and by governments and investors who must fulfil promises made at COP (Conference of the Parties) summits, major businesses are facing ever more stringent sustainability reporting requirements. Big energy users, such as data centers, are in the firing line.

Much of the reporting requirements, and proposed methods of reducing carbon emissions, are going to be complicated and may appear contradictory and counterproductive. For this reason, we can safely predict that many managers will be bewildered and frustrated with the requirements to reduce and report greenhouse gas emissions and other environmental impacts in the years ahead. There are likely to be disagreements between the various parties operating in (and / or overseeing) the data center industry.



There are likely to be disagreements between the various parties operating in — and / or overseeing — the data center industry.

Most managers in the industry recognize the need for action, but trust in regulators is low. Asked if the data center sector needs greater regulation to improve its environmental sustainability, 63% of respondents to a recent Uptime Institute survey (*Uptime Institute Climate Change Survey 2021*) said "yes." But when asked how knowledgeable or informed regulators are in guiding and setting the rules for data center sustainability, respondents to this same survey gave regulators a low competency score.

To date, most of the commitments on climate change made by the industry have been voluntary. This has allowed a certain laxity in the definitions, targets and terminology used — and in the level of scrutiny applied. But these are all set to be tested: reporting requirements will increasingly become mandatory, either by law or because of commercial pressures. Failure to publish data or meet targets will carry penalties or have other negative consequences.

The European Union (EU) is the flag bearer in what is likely to be a wave of legislation spreading around the world. Its much-strengthened Energy Efficiency Directive, part of its "fit for 55" initiative (a legislative package to help meet the target of a 55% reduction in carbon emissions by 2030), is but one example. This legislation will require much more granular and open reporting, with even smaller-sized data centers (around 300–400 kilowatt total load) likely to face public audits for energy efficiency.

For operators in each part of the critical digital infrastructure sector, there may be some difficult decisions and trade-offs to make. Cloud companies, enterprises and colocation companies all want to halt climate change, but each has its own perspective and interests to protect.

Cloud suppliers and some of the bigger colocation providers, for example, are lobbying against some of the EU's proposed reporting rules. Most of these organizations are already highly energy efficient and, by using matching and offsets, claim a very high degree of renewable use. Almost all also publish power usage effectiveness (PUE) data, and some

© 2022 UPTIME INSTITUTE | ALL RIGHTS RESERVED

produce high-level carbon calculators for clients. Significant, step-change improvements would be complex and costly. Additionally, they argue, a bigger part of the sector's energy waste takes place in smaller data centers, which may not have to fully report their energy use or carbon emissions — and may not be audited.

Cloud suppliers are wary about revealing site-level energy use, water use, or carbon emissions related to particular applications. They do not want analysts trying to calculate their levels of utilization and their competitiveness or the energy efficiency of availability zones (rather than of single sites) that involve a high degree of energy-hungry replication. In one recent US case, Google even argued that water use at one of its sites is a "trade secret."

Colocation companies have a particular conundrum. Their energy consumption is high profile and huge; clients now expect colocation companies to use electricity from low-carbon or renewable sources. This requires the purchase of evermore expensive RECs (renewable energy certificates) and / or expensive, risky PPAs (power purchase agreements).

Purchasing carbon offsets or sourcing renewable power alone, however, is not likely to be enough in the years ahead. Regulators and investors will increasingly want to see annual improvements in energy efficiency or in reductions in energy use and carbon emissions. Indeed, any organization that has enrolled in the Science Based Targets initiative (SBTi) has made a commitment to reduce carbon emissions every year.

For a colocation provider, achieving significant energy efficiency gains every year may not be possible. More than 70% of their energy use is tied to (and controlled by) their IT customers — many of these customers are also pushing for more resiliency, which usually uses more energy. This can also apply to bare metal cloud customers.

In most data centers, the IT systems consume the most power and are operated wastefully. To encourage more energy efficiency at colocation sites, it makes sense for enterprises to take direct, Scope 2 responsibility for the carbon in the purchased electricity powering their systems. At present, enterprises in a colocation site are most likely to categorize the carbon associated with the IT as embedded Scope 3, which has weaker oversight and is not usually covered by expensive carbon offsets.

But while many (including Uptime Institute) advocate that IT owners and operators take Scope 2 responsibility, it is clearly problematic. The owners and operators of the IT would have to be accountable for the carbon emissions resulting from poor energy purchases made by their colocation or cloud companies — something many will not yet be ready to do. And, if they are responsible for the carbon emissions, they may have to also take on more responsibility for the expensive RECs and PPAs. But the change might, at least, encourage operators to take on the considerable task of improving IT efficiency.

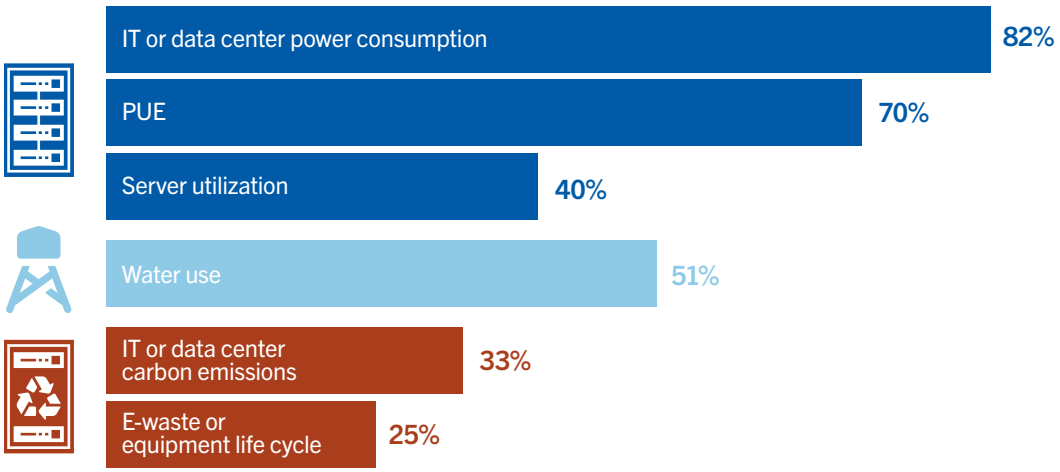
IT energy waste is a challenge for most in the digital critical infrastructure sector. After a decade of trying, the industry has yet to settle on metrics for measuring IT efficiency, although there are good measurements available for utilization and server efficiency

© 2022 UPTIME INSTITUTE | ALL RIGHTS RESERVED

(see **Figure 2**). In 2022, this challenge will rise up the agenda again, as stakeholders once again seek to define and apply the elusive metric of "useful work per watt" of IT. There won't be any early resolution, though: these metrics are specific to each application, limiting their usefulness to regulators or overseers — and executives may fear the results will be alarmingly revealing.

**Figure 2** **Power consumption and PUE are top sustainability metrics tracked**

Which IT or data center metrics do you compile and report for corporate sustainability purposes? Choose all that apply (top respondents only). (n=539)




UPTIME INSTITUTE GLOBAL SURVEY OF IT AND DATA CENTER MANAGERS 2021

UptimeInstitute | INTELLIGENCE

**Stricter environmental policies will force division among industry stakeholders. Faced with growing regulatory and commercial pressure, operators will become frustrated with how to assess and report on sustainability.**

© 2022 **UPTIME INSTITUTE** | ALL RIGHTS RESERVED

# 3. Data center operators ponder the nuclear option

Data center owners and operators worldwide have long been under pressure to use sustainable power — with little to no carbon emissions. Despite a decade of effort, few data center operators are yet to lay claim to having even one data center that uses entirely carbon-free energy.

This is the struggle that has been faced by all businesses worldwide. Both renewable energy generation and emission reduction pledges at the 2021 COP26 summit in Glasgow, Scotland, fell well short of what is required to keep global warming below two degrees Celsius. In 2021, power generation from coal jumped to meet a rising demand for energy — with associated carbon emissions increasing by 3.5%, the International Energy Agency projected.

As major businesses feel a growing sense of urgency to dramatically cut carbon emissions, opinions are starting to shift in favor of nuclear power, which is not classed as clean, but is a near-zero carbon energy source. The digital infrastructure industry, a major global consumer of energy, has a role to play in rehabilitating nuclear, and in marrying it to intermittent renewable energy to provide firm zero-emission power.

There is considerable reluctance to use, or endorse, the use of nuclear power, largely stemming from a fear of meltdowns and concerns about nuclear waste. These worries are likely to be overcome by the need for dispatchable, low-carbon power generation that does not depend on local weather. From 2022, we expect some major data center operators, and industry influencers and leaders, to support nuclear power more actively and openly — even pressuring governments and utilities to invest in this option.

From 2022, we expect some major data center operators, and industry influencers and leaders, to support nuclear power more actively and openly — even pressuring governments and utilities to invest in this option.

We see three areas where the data center industry will (or may) have an active role in supporting a significant shift toward nuclear power. Two of these areas are important and immediate; the third, involving the development of new nuclear technologies, is more speculative.

First, some data center operators are likely to publicly endorse nuclear as a carbon-neutral power source — for some, this is a significant policy change. The EU's Carbon Neutral Data Centre Pact, an industry initiative, has publicly declared nuclear power acceptable in a zero-emission energy mix. US hyperscalers AWS, Google and Microsoft all mention on their websites the role of nuclear power in providing low-carbon electricity.

Second, we expect more data center operators to buy nuclear power, where available, as part of the mix in their PPAs, as a carbon-friendly alternative to 100% renewable energy (over the short term). Historically, most corporate PPAs and green financing mechanisms have been limited to renewable power (and excluded nuclear) to satisfy environmental concerns. In deregulated markets in the US, this has seriously damaged the nuclear generation industry, both financially and it's not their position in terms of energy policy. Some plants are becoming financially unviable, and others are at risk of premature decommissioning.

© 2022 UPTIME INSTITUTE | ALL RIGHTS RESERVED

The financial case for including nuclear power in a PPA contract is that there will be greater long-term price stability – compared with all-renewable PPAs, which can become increasingly oversubscribed, pushing up prices. This long-term price trend, i.e., high and less stable prices, also applies to energy certificates and carbon offsets.

By complementing renewable energy, nuclear also helps data centers achieve truly low-carbon emissions, not just in carbon accounting terms. This is because nuclear can provide the "firm" low-carbon energy that helps to ensure the stability of the grid when renewable energy supply is low.

In general, we expect most power policies aiming for 100% renewable energy to become less tenable (with some geographical exceptions), as skepticism around them grows. Some operators, for example, have started exploring a combination of renewable and nuclear electricity purchases to guarantee a facility is always supplied with low-carbon content.

Without investments in more new plants, nuclear power will gradually diminish in the US, Europe and Japan — even if other countries, such as China and India, continue to invest in the construction of new nuclear plants. Utilities will dispatch coal and gas power to fill the shortfall when there is insufficient renewable energy.

This leads on to the third point. To change this dynamic, some significant private backers, and several governments, including the US, UK and France, are funding the development of new, modularized designs (known as SMRs, or small modular reactors). In a plan revised in 2021, the US Department of Energy (DOE) has committed $1.4B to the construction of a power plant using a cluster of SMRs from NuScale, a US startup. Other SMR developers, such as TerraPower and X-Energy, have also received DOE funding to demonstrate their nuclear reactors. In the UK, the government awarded a total of £210M ($280M) to Rolls-Royce, matched by a private equity investment of £195M ($260M), to move the company's SMR design toward completion. South Korea is close to fully certifying an SMR from local startup the Korea Atomic Energy Research Institute (KAERI). These are only some examples of active development projects in SMR development.

There are three key innovations behind SMRs:

- Simplified reactor and plant design.
- Extensive prefabrication.
- Passive safety systems, which don't rely on active controls for cooling but use convection, pressure or other forces. This eliminates the requirement for emergency backup power, as already agreed by nuclear safety regulators in the case of NuScale.

Not only are SMRs proposed for the grid, but some of the hyperscalers have reportedly expressed interest in funding new sites that can supply power directly. At 10–100 megawatts (MW) electrical output, some of the next-generation modular reactors will be small enough for large data center operators to fund through PPAs. The projected cost for SMR plants, once in volume production, is a few hundred million dollars, and each would take only three to five years to build (compared with the current timeline of seven to 10 years for conventional nuclear plants).

© 2022 UPTIME INSTITUTE | ALL RIGHTS RESERVED

If they deliver, SMRs could potentially become attractive for large data center campuses (potentially connecting directly), and for municipalities with a large concentration of data centers and other large energy users. SMRs also tend to have fuel preloaded for over 10 years at full capacity, eliminating the need for costly, complex and disruptive refueling. There are, of course, many issues and objections, including the need for power redundancy, cost, security, safety and the difficulty of matching the SMR lifespans to the data center.

Opponents argue there is neither the time nor need for next-generation nuclear. But a renewable energy-only strategy is a bet on grid-scale energy storage technologies that do not exist, and on high-capacity power transmission links for which there are no plans. Worse still, a renewable energy-only strategy adds risk to power security due to the unpredictability of extreme weather events because of climate change.

Owing to its more dynamic build activity, and large energy requirement compared with other industries, the data center sector could play an outsized role in rehabilitating nuclear and steering the future of energy.

**The data center industry is dropping its aversion to nuclear energy. Operators will openly promote nuclear as a low-emission power source, incorporate it into their energy purchase portfolios and encourage the development of its new reactors.**

© 2022 UPTIME INSTITUTE | ALL RIGHTS RESERVED

# 4. Concerns over cloud concentration risk grow

While a public cloud provides a flexible, stable and distributed IT environment, there are growing concerns around its use.

The demand for cloud computing was already experiencing rapid growth, but this growth accelerated during the pandemic as more businesses and consumers became increasingly dependent on an ever-expanding online services. While a public cloud provides a flexible, stable and distributed IT environment, there are growing concerns around its use.

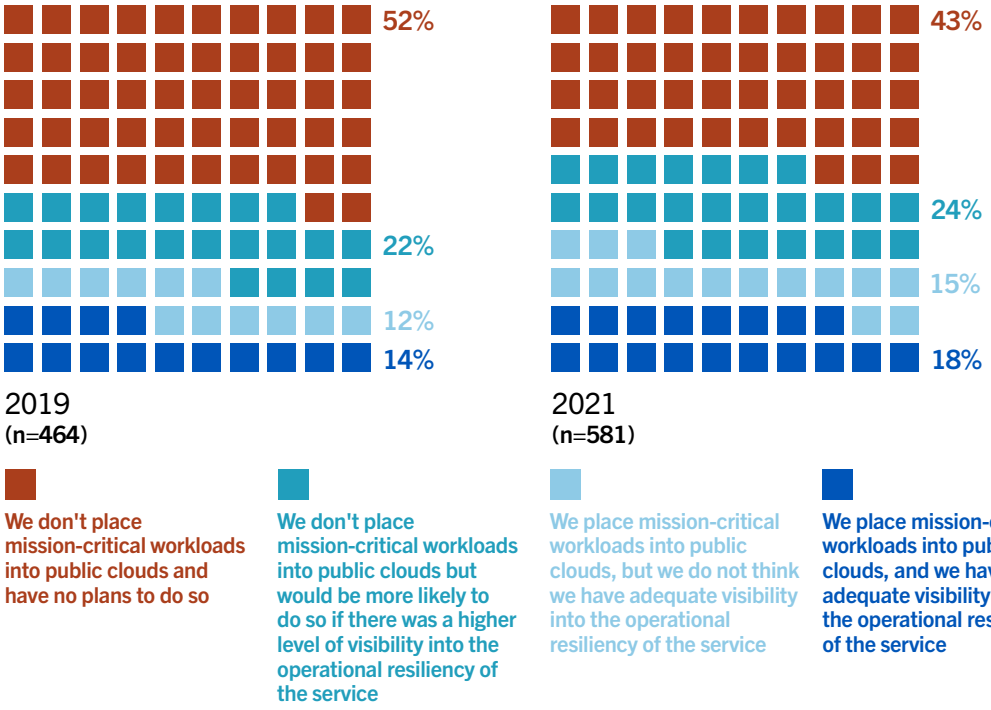
These issues center on resiliency, an area almost all public cloud providers can fairly claim to excel. Organizations with mission-critical workloads build in redundancy at every level of the IT and have processes to work around failures when they do occur. Following some recent high-profile cloud failures, and with regulators asking more questions, there is growing concern that using a big cloud provider is a single point of failure, not just technically but also from a business-risk perspective.

Many organizations and regulators are concerned with the lack of transparency of cloud providers, and the lack of control (see **Figure 3**) that important clients have — some of which are part of the national critical infrastructure. Concentration risk, where key services are dependent on one or a few key suppliers, is a particular concern.

**Figure 3** **More mission-critical workloads in public clouds, but visibility issues persist**

Does your organization have adequate visibility into the resiliency of public cloud operations (e.g., AWS, Azure, Google Cloud Platform) in terms of architecture, availability record, management processes, and full understanding of options?

Cloud strategies are often described as an operating model rather than a location.




UPTIME INSTITUTE GLOBAL SURVEY OF IT AND DATA CENTER MANAGERS 2019, 2021

Uptime Institute | INTELLIGENCE

© 2022 UPTIME INSTITUTE | ALL RIGHTS RESERVED

However, the range, scope of services, management tools and developer environments vary among major cloud providers, often forcing organizations to choose a single provider (at least, for each business function). Even in highly regulated and critical sectors, such as financial services, a multicloud strategy is often neither feasible, nor is it easy to change suppliers — whatever the reason.

In 2021, for example, two major US financial firms Bank of America and Morgan Stanley announced they would standardize on a primary public cloud provider (IBM and Microsoft Azure, respectively). Spreading workloads across multiple clouds that use different technologies, and retraining developers or hiring a range of specialists, had proved too complex and costly.

Unsurprisingly, big cloud providers say that running workloads just in their environment does not lead to an over-reliance. They recommend diversifying within a single cloud to mitigate risk, such as deploying workloads in a cloud using platform as a service (PaaS) and using an infrastructure as a service (IaaS) configuration for disaster recovery. Big providers also point to the distributed nature of cloud computing, which, combined with good monitoring and automated recovery, makes it highly reliable.

Reliability and resiliency, however, are two different things. High reliability suggests there will be few outages and limited downtime, while high resilience means that a system is not only less likely to fail but it, and other systems that depend on it, can quickly recover when there is a failure. And in cloud computing, as in all systems, failures do happen. While in enterprise and colocation data centers, and in corporate IT, the designs can be scrutinized, single points of failure eliminated, and the processes for system failure rehearsed, in cloud services it is mostly (or partly) a black box. These processes are conducted by the cloud provider, behind the scenes and for the benefit of all their clients, and not to ensure the best outcomes for just a few.

Uptime Institute research shows that cloud providers have high levels of reliability, possibly higher than many IT environments, but they are not immune to failure. Complex backup regimes and availability zones, supported by load and traffic management, improve the resiliency and responsiveness of cloud providers, but they also come with their own problems. When issues do occur, many customers are often affected immediately, and recovery can be complex. In 2020, Uptime Institute recorded 21 cloud / internet giant outages that had significant financial or other negative consequences (see *Annual Outage Analysis 2021*).

Mindful of these risks, US financial giant JPMorgan, for example, is among few in its sector taking a multicloud approach. JPMorgan managers have cited concerns over a lack of control with a single provider and, in the case of a major outage, the complexity and the time needed to migrate to another provider and back again.

Regulators are also concerned — especially in the financial services industry. In the UK, the Bank of England is introducing new rules to ensure better management oversight over large banks' reliance on cloud as part of a broader risk-reduction initiative for financial services. And the European Banking Authority mandates that a cloud (or other third-party) operator allows site inspections of data centers. These rules force financial institutions to conduct due diligence on cloud providers before using their services.

© 2022 **UPTIME INSTITUTE** | ALL RIGHTS RESERVED

A newer proposed EU law has wider implications: the Digital Operational Resiliency Act (DORA) puts cloud providers under financial regulators' purview for the first time. Expected to pass in 2022, cloud providers — among other suppliers — could face large fines if the loss of their services causes disruption in the financial services industry. European governments have also expressed political concerns over growing reliance on non-European providers.

In 2022, we expect these "concentration risk" concerns to rise up more managers' agendas. In anticipation, some service providers plan to focus more on enabling multicloud configurations. VMware, for example, freshly spun out of Dell Technologies, is planning to offer tools and services to help managers use multiple clouds, addressing security issues and other cross-platform requirements.

However, the concentration risk goes beyond cloud computing: problems at one or more big suppliers have been shown to cause technical issues for completely unrelated services. In 2021, for example, a technical problem at the content distribution network (CDN) provider Fastly led to global internet disruption; while an outage at the CDN provider Akamai took down access to cloud services from AWS and IBM (as well as online services for many banks and other companies). Each incident points to a broader issue: the concentration of control over core internet infrastructure services in relatively few major providers.

Because these providers have proprietary services that generally don't interoperate, it is often too costly or onerous to use alternative services. But even when customers do plan to do this, the failure of one can shift all the work onto a rival service, overwhelming it.

In October 2021, for example, a cascading issue with Facebook's network caused a global slowdown of the internet, affecting many cloud and IT services, as well as some cellular networks.

How will these concerns play out? Large customers of cloud computing are demanding a better view of cloud suppliers' infrastructure and a better understanding of potential vulnerabilities. As our research shows, more IT and data center managers would consider moving more of their mission-critical workloads into public clouds if visibility of the operational resiliency of the service improves.

While public cloud data centers may have adequate risk profiles for most mission-critical enterprise workloads already, details about the infrastructure and its risks will increasingly be inadequate for regulators or auditors. And legislation, such as the proposed DORA, with penalties for outages that go far beyond service level agreements, are likely to spur greater regulatory attention in more regions and across more mission-critical sectors.

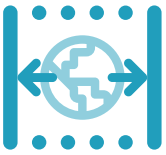
**Control over critical infrastructure is increasingly in the hands of a small number of major providers, raising questions around resiliency. Customers and regulators will demand more transparency into providers' infrastructure and vulnerabilities, and many will explore multicloud strategies.**

© 2022 UPTIME INSTITUTE | ALL RIGHTS RESERVED

# 5. Supply problems favor standardization and scale

The COVID-19 pandemic has stressed supply chains globally, resulting in higher prices and increased delivery times for many components. Shortages and delays are likely to persist into 2022 and beyond as demand for IT, and for new data center capacity, continues to increase.

To keep pace, data center owners and their suppliers are taking steps to mitigate supply-chain disruption — but the playing field is far from level. The biggest operators, and especially those with a multinational operation, are using their buying-power, relationships and scale to secure supplies, while smaller ones may struggle. This will further widen the cost advantage that bigger data center builders and operators already enjoy.



The biggest operators, and especially those with a multinational operation, are using their buying power, relationships and scale to secure supplies, while smaller ones may struggle.

Since the start of the pandemic, delivery times for certain critical data center equipment, including intelligent breakers, cooling components and uninterruptible power supply (UPS) systems, have increased from one to six months in many regions. Delays are often due to a shortage of parts, such as semiconductors, compressors or fans. Supply chain disturbances have led to higher prices for silicon and other raw materials (including copper), as well as for transportation of components by both sea and land.

Despite these issues, there remains an overall expectation that the total cost and time to deliver new data center capacity (per MW of IT) will continue to decrease, as it has done for many years (see the Uptime Institute report *Best-in-class data center provisioning*).

To meet expectations for faster, more cost-efficient builds, data center operators are collaborating with their key suppliers more closely, including with general contractors and system / component vendors. The biggest cloud and internet giants have been working with suppliers to value engineer products and components for many years, but this collaboration now includes component availability. This might mean, for example, densification (raising the IT load per rack to reduce the facility footprint and cost) and increasing the use of off-site system integration, such as making use of electrical power skids, containerized cooling systems and other prefabricated components.

Greater collaboration can also mean more joint engineering between the owner and suppliers at the design phase. This will ensure available components are being used and, in some cases, enable compromises on specifications. Prefabrication has been shown to shorten delivery times and costs for large-scale data center builds when multiples of identical units are used (speeding up both construction and commissioning). More extensive use of industrialization has benefited the entire data center sector, as repeated use of the same designs encourages suppliers to continuously invest in more efficient processes for construction and commissioning. Many smaller data centers, however, continue to ask for custom one-off designs, which in turn exposes them to supply chain difficulties and increased costs.

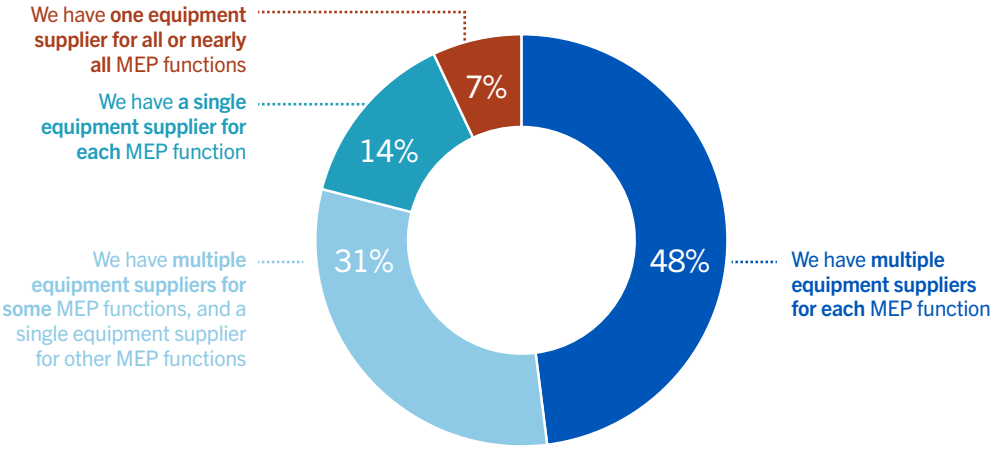
© 2022 UPTIME INSTITUTE | ALL RIGHTS RESERVED

Across the sector, research shows that data center operators are using various tactics to mitigate the impact of supply chain shortages. About a third of operators (32%) surveyed added a second (or more) source supplier to cope with supply disturbances during the pandemic, while 36% added inventory for this purpose.

Even so, the sector clearly remains vulnerable to sourcing problems and to reliance on a single supplier (see also **Concerns over cloud concentration risk grows**), as shown in the *Uptime Institute Supply Chain Survey 2021*. More than half of the respondents to this survey are using a single equipment supplier for one or more mechanical, electrical and plumbing function; and 37% lack adequate visibility into the supply chains of their key vendors. This leaves much of the data center sector exposed to vendor-specific delays and cost increases (**Figure 4**).

**Figure 4** **Half of data centers are exposed to vendor-specific delays**

Which statement best describes the sourcing strategy of your data center's capital equipment suppliers for mechanical, electrical, and plumbing (MEP) functions? (n=226)




UPTIME INSTITUTE SUPPLY CHAIN SURVEY 2021 Uptime Institute | INTELLIGENCE

Operators with multiple large facilities typically have several data center projects running in parallel and are best positioned to use strategic buying tactics. This includes combining their own forecasts of requirements with supplier monitoring to identify potential supply issues as early as possible. One major systems supplier takes measures to ensure that its own suppliers have adequate suppliers by using its commercial influence to reach further up the supply chain.

Staffing is another area where supply is not meeting demand. Uptime Intelligence research shows that almost half of data center owner / operators (47% of those surveyed) have problems finding candidates for open positions. About a third (32%) have difficulty retaining staff as they are being poached mainly by competitors and by companies outside the data center sector. This will continue in 2022 and beyond, with those companies able to pay larger salaries being the ones most able to attract and retain staff.

© 2022 UPTIME INSTITUTE | ALL RIGHTS RESERVED

While supply chain disruptions are likely to continue during 2022 (resulting in higher costs and longer delivery times), it is the large operators that are likely to maintain low overall project costs and tight timeframes. These large operators will need to do more planning, more coordination and more collaboration across the supply chain, and more upfront work to complete projects. Most of these operators, however, will be able to influence and absorb these changes. Without adapting to more industrialized, standardized building blocks, smaller data center operators are likely to fall behind and the division between large and low-cost capacity and the rest of the market is likely to widen.

**Interruptions to supply chains are affecting all suppliers, but larger scale data center operators are positioned to benefit by leveraging their buying power and influence. Smaller operators that embrace a standardization approach are likely to have the best chance of remaining competitive.**

© 2022 **UPTIME INSTITUTE** | ALL RIGHTS RESERVED

Appendix

# Five data center trends for 2021

## 1. Accountability — the "new" imperative

Enterprises want more cloud and greater agility, but they can't outsource responsibility — for incidents, outages, security breaches or even, in the years ahead, carbon emissions. In 2021, hybrid IT, with workloads running in both on- and off-premises data centers, will continue to dominate, but investments will increasingly be constrained and shaped by the need for more transparency, oversight and accountability. More will be spent on cloud and other services, as well as in on-premises data centers.

## 2. Smarter, darker data centers

Following a scramble to effectively staff data centers during a pandemic, many wary managers are beginning to see remote monitoring and automation systems in a more positive light, including those driven by artificial intelligence (AI). An adoption cycle that has been slow and cautious will accelerate. But it will take more than just investment in software and services before the technology reduces staffing requirements.

## 3. Edge — the next frontier

Significant new demand for edge computing, fueled by technologies such as 5G, the internet of things (IoT) and AI, is likely to build slowly but the infrastructure preparation is underway. Expect new alliances and investments across enterprise, mobile and wireline networks, and for a wide range of edge data centers, small and large. Smart and automated software-defined networks and interconnections will become as important as the physical infrastructure.

## 4. Sustainability: More challenging, more transparent

For years, operators could claim environmental advances based on small, incremental and relatively inexpensive steps — or by adopting new technologies that would pay for themselves anyway. But the time of easy wins and greenwashing is ending: Regulators, watchdogs, customers and others will increasingly expect operators of digital infrastructure to provide hard and detailed evidence of carbon reductions, water savings and significant power savings — all while maintaining, if not improving, resiliency.

## 5. A surge of innovation

Data center operators (and enterprise IT) are mostly cautious, if not late, adopters of new technologies. Few beyond hyperscale operators can claim to have gained a competitive advantage through technology. However, several new technologies are maturing at the same time, promising advances in the performance and manageability of data centers and IT. Storage-class memory, silicon photonics, ARM servers and software-defined power are ready for greater adoption.

© 2022 UPTIME INSTITUTE | ALL RIGHTS RESERVED

# About the authors

## Andy Lawrence

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**

## Rhonda Ascierto

Rhonda Ascierto is Uptime Institute's Vice President of Research. She has spent 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**

## Daniel Bizo

Daniel Bizo is Uptime Institute's Research Director. He has been covering the business and technology of enterprise IT and infrastructure in various roles for 15 years, including a decade as an industry analyst and advisor.

**Contact: dbizo@uptimeinstitute.com**

## Tomas Rahkonen

Tomas Rahkonen is Uptime Institute's Research Director of Distributed Data Centers. He has extensive experience from design and delivery of data centers to four continents and from mobile telecommunications networks and applications.

**Contact: trahkonen@uptimeinstitute.com**

### About Uptime Institute

Uptime Institute and the company's Tier Standard is the globally recognized digital infrastructure authority known for the creation and administration of the world's most adopted standards for data center performance and resilience. For over 25 years, Uptime Institute has been providing customers with the assurance that their digital infrastructure can perform at a level consistent with their business needs, across a wide array of operating conditions. With its data center Tier Standard & Certifications, Management & Operations reviews, Digital Infrastructure Resiliency Assessments and other services, along with our accredited educational curriculum for data center professionals, Uptime Institute helps organizations optimize critical IT assets while managing costs, resources, and efficiency. Today, thousands of companies rely on Uptime Institute to enable their digital-centric business success.

Uptime Institute — The Global Digital Infrastructure Authority®, a division of The 451 Group, has office locations throughout the world, including North and South America, Europe, UAE, Russia, and Asia.

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**

© 2022 **UPTIME INSTITUTE** | ALL RIGHTS RESERVED