451 RESEARCH'S
2019 PREVIEW REPORTS

Analysis of the technologies and trends that will drive and reshape enterprise technology markets in the year to come

451 Research®

DEC 2018

2019 Trends in Datacenter Services & Infrastructure

Kelly Morgan, Research Vice President, Services

Dan Thompson, Research Director - MTDC

Daniel Bizo, Senior Analyst, Datacenters and Critical Infrastructure

Penny Jones, Research Director - MTDC & Managed Services

Datacenters are under pressure: They must be the right size, in the right location, with the right connectivity. Despite efforts to move redundancy into applications, datacenter infrastructure is expected to function 24/7, as efficiently as possible. The largest customers are pushing for innovation and faster builds. Datacenter vendors and service providers are responding by becoming more flexible and adopting new technologies, but more change – particularly from the growth of IoT and 5G deployments – will be required for the industry in the years ahead.

About the Authors

Kelly Morgan

Research Vice President, Services

Kelly joined 451 Research in April 2011 as an analyst covering the economics and finances of the datacenter and hosting industries. In her current role, she leads the services analyst teams, which track hosting, managed services and multi-tenant datacenter providers worldwide. Coverage includes providers, market size, datacenter supply and demand, and the dynamics of the industry overall.

Dan Thompson

Research Director - MTDC

Daniel Bizo

Senior Analyst, Datacenters and Critical Infrastructure

Penny Jones

Research Director - MTDC & Managed Services

Key Findings

- Cloud providers will continue to drive demand for large datacenters in low-cost locations, as well as for network-dense connectivity facilities. However, the rise of 5G and IoT will also push demand for smaller edge datacenter facilities that require different business models and automation. Datacenter providers will need to target their customers carefully and understand exactly what qualities and services customers seek.

- Data analytics, cloud, prefabrication, multi-site resiliency and open-source architectures are being applied to datacenters. This is creating intelligent, connected and purpose-built critical infrastructure that is responding more rapidly to pressure from customers and technological change.

- Datacenter operators will need to plan for environmental change: warmer temperatures outside, changes in water availability, customer and regulatory requirements around power sources, and greater natural disaster risk. Engineering based on historical worst-case climatic conditions might not be enough in the future. Customers will increasingly ask for a broader vision around environmental responsibility.

Executive Summary

Introduction

Things are changing quickly for the datacenter industry. Cloud adoption and the rise of hyperscale IT and cloud firms has driven demand, particularly for large-scale space, but also for innovation. From new chips to Open Compute to liquid cooling to solid-state storage, large IT firms have been pushing for new IT infrastructure approaches, often in collaboration with vendors. Changes in systems architecture and IT components also impact datacenter design. Innovations that combine and optimize IT at the physical, logical and virtual levels are driving new opportunities – and new challenges – for operators and owners of datacenters and critical infrastructure as well as for their suppliers.

"Providers that can offer automation, scalability and flexibility as well as customization will stand out from the crowd in the short term while possibly being the only ones to survive in the long term."

Datacenter providers are increasingly building in locations and at a scale that will appeal to large IT and cloud firms. However, there will likely also be demand in the future for smaller, edge datacenters related to 5G and IoT deployments. This could lead to the growth of facilities that are not 'traditional' datacenters. Some, such as smart buildings and factories, may support megawatts of IT capacity, be heavily connected, and require multiple power distribution and storage systems for energy efficiency and availability. Others, such as distributed micro-datacenters (e.g., at cell towers), may support less than 100kW of capacity but be grouped together under centralized software control to simplify management. Regardless of datacenter size, operating facilities efficiently and continuously is becoming increasingly complex. This is driving a need for software and analytics-driven services to help monitor and manage these sites.

All of this is required to deliver what 451 Research calls Invisible Infrastructure. This is defined in our *451 4SIGHT report* as infrastructure that 'just works': instantly available, operating and scaling regardless of specific requirements, and billed and metered in a flexible manner. Customers increasingly expect 'cloud-like' consumption options for all of their infrastructure needs, and providers that can offer automation, scalability and flexibility as well as customization will stand out from the crowd in the short term while possibly being the only ones to survive in the long term.

451 Research's 2019 Trends in Datacenter Services & Infrastructure

Source: 451 Research, 2018

	Winners	Losers
Providers Will Accommodate or Cooperate With the Cloud	Providers that can attract multiple hyperscale deployments to build a sticky cloud ecosystem; those that can differentiate with distinction; enterprises moving to hybrid IT	Providers chasing hyperscale deployments to the exclusion of other opportunities; providers stuck in 'pure-play' thinking; those not open to partnerships; regional cloud service providers only selling IaaS
IoT and 5G Will Change the Way We Think About the Edge	Providers with datacenters in multiple remote locations, an interconnection platform, or services beyond colocation; telcos that can turn their local central offices into edge datacenters; local development boards seeking to gain foreign direct investment	Local facilities without a lot of connectivity options; providers targeting the content delivery business without sites in edge cities; small local players that only provide colocation; cloud consumers concerned by cost
Datacenters Will Be Increasingly Industrialized, AI-Assisted and (Re) distributed	Operators that are willing to adapt quickly and those that understand the value of software	Operators with entrenched habits and no long-term technology outlook; those overly focused on the endgame
Rising Chip Power Will Force Rack Density Higher, Boosting Interest in Liquid Cooling	Operators willing and able to add direct liquid cooling (DLC) cabinets; enterprises and cloud firms that are able to take advantage of high-power chips	Datacenter operators that do not keep an eye on the DLC trend
Climate Change and Sustainability Will Become Hot Issues	Operators that have already made efforts to boost efficiency and use renewables	Operators that are comfortable with their current disaster plans

Methodology

Reports such as this one represent a holistic perspective on key emerging markets in the enterprise IT space. These markets evolve quickly, though, so 451 Research offers additional services that provide critical marketplace updates. These updated reports and perspectives are presented on a daily basis via the company's core intelligence service, 451 Research Market Insight. Forward-looking M&A analysis and perspectives on strategic acquisitions and the liquidity environment for technology companies are also updated regularly via Market Insight, which is backed by the industry-leading 451 Research M&A KnowledgeBase.

Emerging technologies and markets are covered in 451 Research channels including Applied Infrastructure & DevOps; Cloud Transformation; Customer Experience & Commerce; Data, AI & Analytics; Datacenter Services & Infrastructure; Information Security; Internet of Things; Managed Services & Hosting; and Workforce Productivity & Compliance.

Beyond that, 451 Research has a robust set of quantitative insights covered in products such as Voice of the Enterprise, Voice of the Connected User Landscape, Voice of the Service Provider, Cloud Price Index, Market Monitor, the M&A KnowledgeBase and the Datacenter KnowledgeBase.

All of these 451 Research services, which are accessible via the web, provide critical and timely analysis specifically focused on the business of enterprise IT innovation.

For more information about 451 Research, please go to: *www.451research.com*.

451 Research's 4SIGHT: Empowering the Digital Revolution

Throughout this report, you will see the below graphic indicating which of our *451 4SIGHT* themes each trend relates to. 4SIGHT is 451 Research's look into the future of information technology, organized around four main themes that we expect to shape the digital transformation agenda over the next 10 years and beyond. The 4SIGHT report is available to all clients via our research dashboard. For more information, *visit our website*.

Table of Contents

Trend 1: Providers Will Accommodate or Cooperate With the Cloud

IMPACT TO THE MARKET

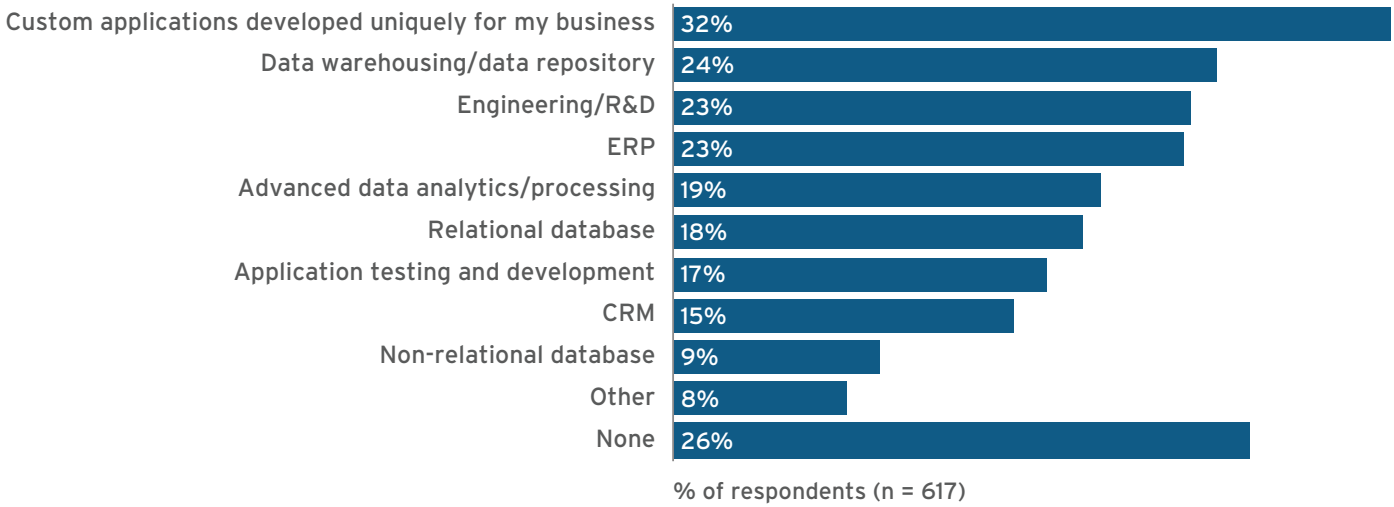
Implication: *In a recent survey we found that 38% of colocation providers were just as likely to see themselves competing with hyperscale cloud firms as partnering with them. However, the products and services offered by datacenter providers are fundamentally different from those that hyperscale cloud players offer. Datacenter firms should focus on how to accommodate cloud providers – either through scale or interconnection options – or cooperate with them by offering hybrid products and migration services.*

The public cloud is a tool and has its uses, but colocation and private cloud have their uses as well. Our data consistently shows that enterprises operate multi-venue deployments, and most plan to continue to mix these in the future. For major cloud providers, however, the old saying, "When all you've got is a hammer, everything looks like a nail" sometimes applies. To those organizations, every workload looks like a potential cloud workload, so that is what much of their messaging has focused on. Our data shows this is not necessarily the case, though it is difficult for enterprises to weigh all options and determine best execution venues when non-hyperscale firms can barely be heard above the cloud din. Hyperscale cloud providers are beginning to realize this, however, and may end up promoting other options. Amazon Web Services (AWS) has begun to offer colocation services on its marketplace, for example. Datacenter providers will increasingly need to communicate the value of the different venues available and offer a helping hand to get organizations where they need to go. They can be the glue linking cloud, connectivity and service options together.

In 451 Research's *Voice of the Enterprise (VotE): Cloud, Hosting & Managed Services, Workloads and Key Projects 2018 survey*, we asked the question: Which, if any, of the following workloads does your organization consider to be non-cloud workloads, meaning currently unsuitable to run in a public cloud (IaaS, PaaS or SaaS) environment? Only 26% of respondents felt all of their applications were suitable for the cloud. Although it may be possible to make any applications work in the public cloud, it can be a difficult and expensive task, so the majority of IT leaders believe that something in their environments would not be a good fit (see Figure 1).

Figure 1: Non-Cloud Workloads

Source: 451 Research's Voice of the Enterprise: Cloud, Hosting & Managed Services, Workloads and Key Projects 2018
Q. Which, if any, of the following specific workloads does your organization consider to be non-cloud workloads, meaning currently unsuitable to run in a public cloud (IaaS, PaaS or SaaS) environment? Please select all that apply.



A deep dive into offering various managed services that support public cloud may not be the best fit for all datacenter providers. However, there remains strong demand from customers for managed services in addition to public cloud services, and they are looking to a variety of firms, from systems integrators to managed services specialists to datacenter operators, to offer these. In addition, while IT organizations have become more comfortable with security in the public cloud, it remains a concern. The challenge for datacenter providers is that security conversations they have had in the past have revolved around physical security. While physical security is important, it is no longer the key concern for enterprises when evaluating workload locations. Datacenter providers need to be able to address logical security concerns in addition to physical security, either through tight provider partnerships or through an in-house delivery of security services (managed or otherwise).

A number of providers are looking at ways to reduce cost-to-build and time-to-build for hyperscale customers. This includes firms such as Equinix that have not traditionally built large-scale wholesale facilities targeted at cloud providers. Equinix has launched a Hyperscale Infrastructure Team (HIT) to offer datacenter supply suited to hyperscale requirements. Many firms, such as CyrusOne and NTT/e-Shelter, are innovating to reduce the time required to build facilities – in some cases to nine months or less. E-shelter, for example, says it has deployed a new 'agile' build model, streamlining construction processes by undertaking all design, procurement and project management up front instead of throughout the datacenter's phased lifespan. This approach means they can secure contractors for the entire job and bring supply online inside the building's shell rapidly.

Recommendations

- **Understand your customers' requirements and focus your business model accordingly.** Trying to provide wholesale datacenter space and managed services makes it difficult to prioritize where to spend capital. Large cloud providers generally need space fast, while in many cases, end users will be looking to offload elements of risk, which are two fundamentally different routes for a datacenter business. Know what your customers are struggling with and be sure you have taken the best approach to offering it.

- **Choose your partners well.** For providers that don't have a managed service arm already, building it from the ground up isn't the only way forward. Many providers will work with third parties to provide elements of datacenter services, from migration to security and networking. Know how your partners like to work and have the correct parameters in place to ensure they are treating customers with the same level of service you promise them under contract. This is a great place to look at acquisitions as well for a more long-term answer.

- **Service providers should formalize (where possible) relationships with key technology vendors and integrators to keep up to speed with new products and trends.** They should dedicate time to partner days and special events and make someone in the organization responsible for generating insights into the latest technology innovations. This can be just as true of network and compute hardware as it is generators and UPS systems. Efficiencies can be won on all fronts.

Winners

- **Providers that can attract multiple hyperscale deployments to build a sticky cloud ecosystem.** Cloud and IT firms have been leasing large amounts of space and power. The worry for providers is that these customers will build their own sites down the road and withdraw from leased space. One possible remedy: Ensure that multiple hyperscale firms plus their partners and customers are connected to your campus.

- **Those that can differentiate with distinction.** The opportunity is not in offering as many services as possible to the end user; it is about knowing the type of customer you are targeting and looking at key pain points. Providers that succeed will be those that start off by doing a few things well, and then expand as the need arises.

- **Enterprises moving to hybrid IT.** Enterprise users will benefit from simplified contracts and easier management of hybrid IT environments if they choose a partner that can provide the right mix of services, especially those looking to outsource IT and focus on core business applications and services.

Losers

- **Providers chasing hyperscale deployments to the exclusion of other opportunities.** Providers are focused on building datacenters for hyperscale providers because they are taking large amounts of space very quickly. However, there continues to be demand from enterprises, systems integrators, service providers and other firms that may want smaller amounts of space to start with but could grow quickly.

- **Providers stuck in 'pure-play' thinking.** There will always be customers that require simple datacenter space or services, but providers that don't understand that enterprises now need to discuss services around the negotiation table will find this market is diminishing as IT environments become more complex.

- **Providers not open to partnerships.** The cloud has made not only flexibility but also choice a key part of the datacenter conversation. Few will be looking to form only one relationship with one vendor moving forward. A provider that can prove it is open to partnering where required and adding services across the stack will be more able to prove it is open to catering to an end user's future needs.

- **Regional cloud service providers only selling IaaS.** While IaaS is growing quickly as a service category in Europe, it is doing so from a very small base, driven by the super-scale providers with the capacity, bandwidth and portfolio to drive the market. Those regional service providers that do not offer additional managed or professional services will find themselves out-competed.

Trend 2: IoT and 5G Will Change the Way We Think About the Edge

IMPACT TO THE MARKET

Implication: *The rollout of 5G and the growth of IoT devices are expected to drive demand for vast amounts of data storage and processing capacity. The question is where this capacity will need to be and who will own it. The answer could be disruptive to the current centralized datacenter approach, but also present opportunities – particularly in offering private cloud to particular verticals such as healthcare and government.*

So far, the 'edge' has meant different things to different people. Some companies define the edge as a device, while others might define it as a cell tower, and others still, a datacenter in a smaller city. The reality is that most providers define the edge as a point just shy of where their services stop, and in today's world, our *VotE: Internet of Things, Workloads and Key Projects 2018* survey data suggests that may not be far enough. In 2018 we asked respondents where they transport IoT data for initial analysis, and then where they transport it – or keep it – for further analysis. Roughly 34% indicate that they send the data to a third-party service provider (a public cloud, a colocation facility or network provider infrastructure) for initial analysis (see Figure 2). A much higher number plan to execute additional analysis and store the data off-site, roughly 67%, but even then, many will also store and process data either on the device or in a company-owned facility (see Figure 3).

Figure 2: IoT Data Initial Storage and Analysis

Source: 451 Research's Voice of the Enterprise: Internet of Things, Workloads and Key Projects 2018
Q. Where is your IoT data initially stored and analyzed?

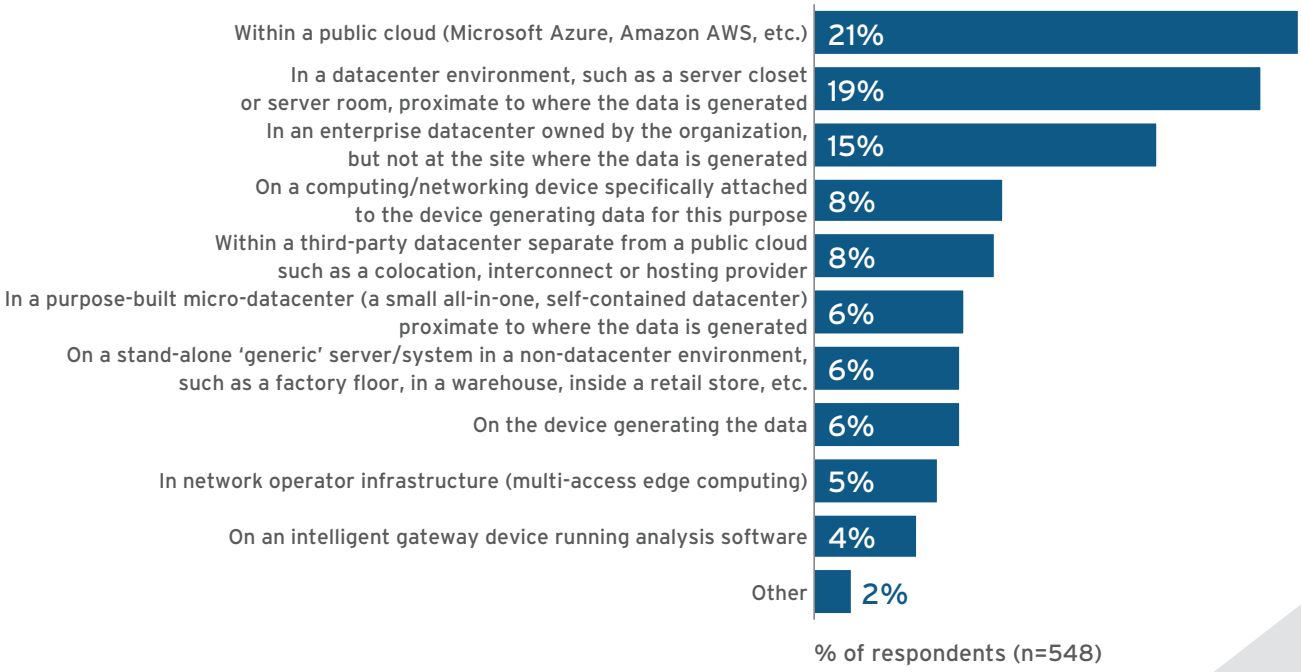
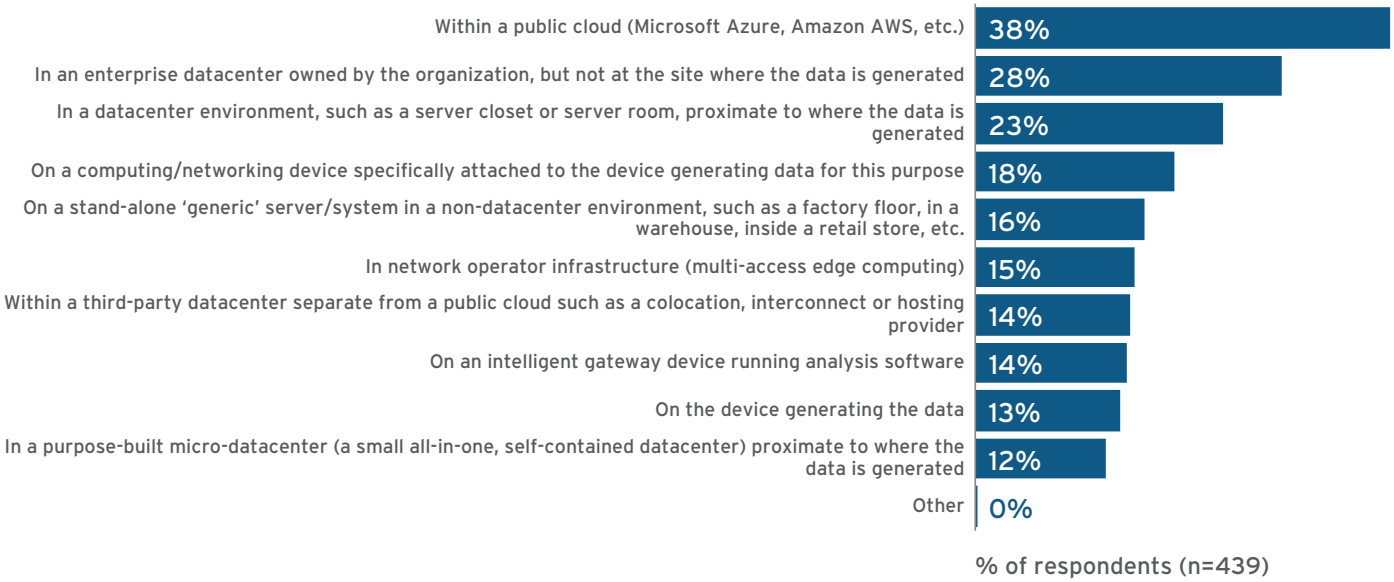


Figure 3: Store IoT Data and Execute Additional Analysis

Source: 451 Research's Voice of the Enterprise: Internet of Things, Workloads and Key Projects 2018
Q. Where do you subsequently store, or plan to store, your IoT data and execute additional analysis? Please select all that apply.

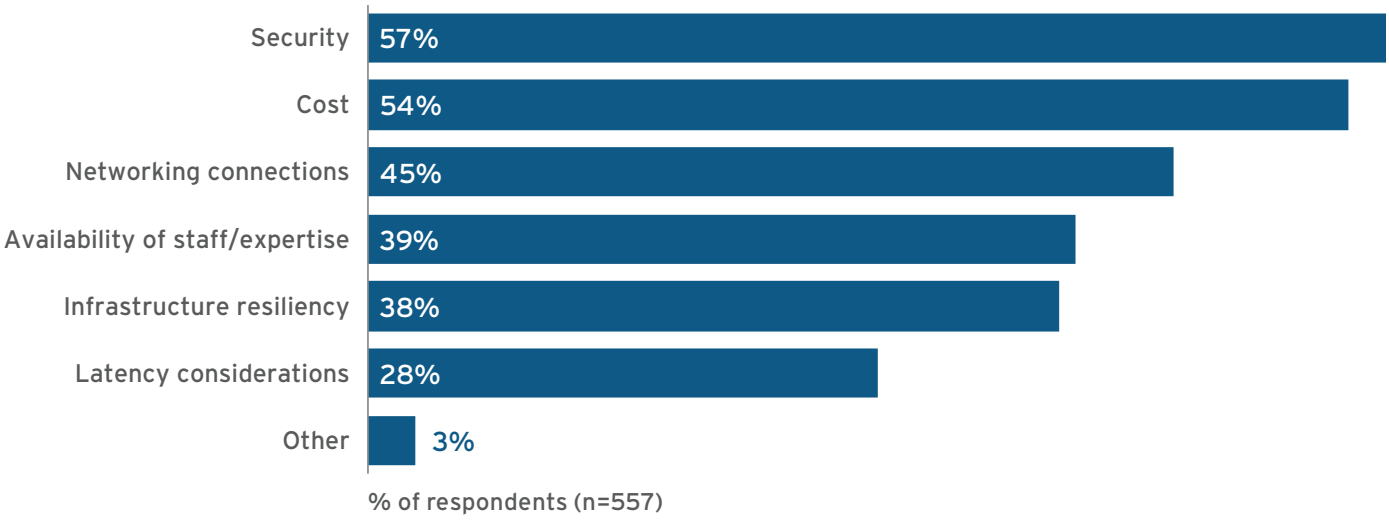


For many industrial IoT devices, latency is of extreme concern because some form of feedback is needed instantly or near instantly (a belt line needs to speed up or slow down, a car needs to apply the brakes, etc.). When the initial analysis must happen at or very near the device, processing the data in a datacenter situated tens or hundreds of miles away is not an option, especially when the devices are in an industrial area with poor connectivity options. For the data sent to company-owned facilities, however, that is a different story. In our 2018 Workloads and Key Projects study, we delved deeper into the placement question, and found that the top three reasons enterprises land the data where they do are the same rebuttals datacenter providers (and frankly cloud providers) have been hearing since the beginning: security, price and connectivity options (57%, 54% and 45%, respectively, as shown in Figure 4).

Figure 4: Determining Best Execution Venue for an IoT Workload

Source: 451 Research's Voice of the Enterprise: Internet of Things, Workloads and Key Projects 2018

Q. In general, which factors are most influential when determining the best execution location or venue for an IoT workload? Please select all that apply.



For datacenter providers that offer dedicated private cloud, an opportunity seemingly exists for a managed private cloud to be deployed on site, with the potential of a larger deployment in a hosted environment inside the provider's datacenter. The 'on-site plus hosted in a third-party facility' hybrid model is one we have not seen widely leveraged yet, and looks to be a good opportunity for disruption in the verticals still going it alone from an infrastructure perspective. Clearly providers need to be able to solve for security and connectivity, while they are already accostomed to answering the price question, even when bundled with a managed service.

The good news is this same line of thinking can be applied to industries like healthcare, where imaging machines and other use cases look very similar to their more industrial application counterparts from an underlying infrastructure perspective. Thinking specifically about the medical imaging application, it is not uncommon for doctors to view x-ray/PET scan/ultrasound results in the same building, or in close proximity to the devices that created the very large datasets associated with those imaging types. To that end, it typically does not make sense to ship those images off to a far-away datacenter because latencies can slow down the experience for the doctors (who are notoriously sensitive to workflow slowdowns). In a similar design as that proposed previously, an on-site managed cloud could be leveraged for immediate viewing and near-term storage, while an off-site hosted platform could manage the long-term storage. Obviously for the medical industry in the US and many other countries, regulations exist as to how this data is to be handled and secured, which frankly represents more opportunities for datacenter/managed service providers with regulatory/compliance knowledge and product sets.

Recommendations

- **Remember that providing edge-ready datacenters is not as simple as building a smaller site in a smaller location.** Edge capacity may require different form factors than past buildouts and connectivity that can handle the new distributed nature of IT. It may also require services beyond colocation to be truly useful to the industries currently leveraging edge and near-edge devices. It is about providing a multi-tier IT service architecture connected by the network.
- **Providers seeking to go down this route must invest in R&D and understand how customers will want to connect,** as well as what parts of the infrastructure those customers are comfortable outsourcing. Enterprises building an edge strategy will need to be provided with controls to ensure data sovereignty and policy setting that can be enacted with a carefully thought-out plan that highlights architecture requirements.
- **Providers building edge datacenters will also need to carefully think about product delivery and pricing models** and identify partners that might be able to help provide additional services – including in-region and in-language support. This can be further bolstered with an understanding of pertinent regulatory requirements and products that can help companies on their compliance journey.

Winners

- **Providers with datacenters in multiple remote locations** that can service customers with a single bill and a single pane of glass for all their facilities.
- **Providers with an interconnection platform** that can offer connectivity to multiple parties in one location.
- **Providers with services beyond colocation** that can tailor an industry-specific offering to those closest to it.
- **Telcos that can turn their local central offices into edge datacenters.** In some cases, this will be difficult due to the quality of the facilities and lack of connectivity options, but there may still be enough demand to make upgrading some of these facilities worthwhile.
- **Local development boards seeking to gain foreign direct investment.** Many hyperscalers are aiming for the second-tier markets in 2019 and eventually a presence in each country.

Losers

- **Local facilities without a lot of connectivity options** that have not seen success with services.
- **Providers targeting the content delivery business** but that do not have sites in multiple 'edge' cities.
- **Small local players that only provide colocation** – similar to wholesale providers but without the scale – and that do not have the density or connectivity to appeal to content delivery and other edge players.
- **Cloud consumers concerned by cost.** As has been witnessed in Germany, building regional or edge datacenters can come at a higher cost (without the efficiency gains large buildouts at scale can bring and with new support or service teams on the ground to meet regulatory requirements).

Trend 3: Datacenters Will Be Increasingly Industrialized, AI-Assisted and (Re)distributed

IMPACT TO THE MARKET

Implication: *Multi-tenant datacenter (MTDC) providers and suppliers alike are forced to make growing use of factory prefabrication of subsystems to streamline construction, improving speed and quality control. Datacenter management as a service (DMaaS) that helps to improve efficiency, prevent outages and lower maintenance costs will likely become a must-have as machine and deep learning models begin to extract more value from telemetric data. As the size and complexity of infrastructure grows, the ability to analyze data not only from facilities but also from IT infrastructure and applications will become a much-needed feature of advanced datacenter infrastructure management (DCIM) software. The closer a facility is to the edge (away from the core), the more essential monitoring, management and rapid deployment become.*

The datacenter industry is on the verge of a resource crunch in its challenge to build and operate a rapidly growing number of modern facilities needed to underpin cloud and enterprise infrastructures. Over a third of operators already report challenges recruiting skilled staff, according to the Uptime Institute's 2018 annual survey (Uptime Institute is an independent operating division of The 451 Group). A similar number have already experienced or expect staffing cuts. This calls for more efficient processes for building and operating datacenters. At the same time, organizations expect datacenters to work consistently while also becoming more efficient. This is not a given, despite decades of experience. Uptime reports outages have actually increased recently: 31% of respondents to its survey say they have suffered failures that impacted their IT service. This may be due to the increased complexity of datacenters and the interdependencies of different systems and different facilities (remembering that most datacenters handle more work each year). There has also been evidence, in earlier research by Uptime, that failures tend to occur most commonly either during periods of technology change and investment or at sites where there has been underinvestment and legacy assets are not upgraded. Many sites fall into one of these two categories, particularly as the move to cloud leads some enterprises to underinvest in their on-premises infrastructure.

Notwithstanding the push toward more distributed resiliency in cloud-native applications, it will be a slow and gradual shift and datacenters in the foreseeable future need to be highly reliable to avoid service outages. At the same time, there is relentless pressure to reduce build costs and bring new capacity online faster. In addition, edge computing would seem to encourage cloud services to be available closer to where consumers and businesses are located to optimize services delivery and meet data governance regulations.

Whatever shape and form it takes, the proliferation of edge computing will strain existing deployment and operational practices that operators have developed around the idea of a relatively small number of centralized datacenter sites. Availability and reliability of connectivity (route diversity, capacity, load-balancing, security, etc.) between all the locations digital services delivery depends on is paramount too. In this context, 451 Research sees prefabricated modular (PFM) datacenters, DMaaS and more robust distributed resiliency schemes becoming indispensable in 2019 and the years that follow.

PFM Datacenters Everywhere

Whether small or large, a turnkey datacenter or an electrical subsystem, prefabrication has gained currency with operators from all walks of life. Moving much of the complexity of datacenter construction into a factory-like environment yields a multitude of benefits, chiefly better quality controls (including pre-commissioning), higher labor productivity and faster delivery of major subsystems (six months and under). The industrialized approach also promotes optimization and improved designs to reduce the cost of materials and improve the manufacturing and testing processes.

Even more so than centralized sites, distributed edge computing facilities (next-generation telecom edge, industrial edge, transport and logistics) will rely on PFM facilities to deliver IT capacity where it is needed. Robust modular rooms, containerized and custom enclosures, and micro-modular datacenters (MMDCs) will all play a role. 451 Research believes MMDCs, an emerging form factor, are particularly well-positioned for the growth of the distributed edge. The concept of micro-facilities and packaged equipment is well understood and widely utilized in other industries, including telecommunications, but the datacenter industry in general lags behind in adoption of industrial techniques despite evidence of their utility. Self-contained, highly integrated and compact, MMDCs are quick and easy to install. 2019 will see a growing stable of products and vendors racing to gain a head start over their competitors.

As the datacenter footprint scales both up (many-megawatt sites) and out (many more small locations), the benefits of PFM methodology will become even more pronounced over traditional on-site construction and fit-out.

Datacenter Management as a Service

DMaaS goes beyond just DCIM as a service. It is part of a long-term evolutionary change toward integrating physical DCIM with other services, including – but not limited to – IT workload management, energy management, connectivity and business costing. DMaaS delivers some of the benefits of DCIM, but with lower barriers, by offering a simplified, low-touch and low-cost (or at least opex-based) alternative.

A key area of focus is monitoring of the health of the datacenter and, over time, condition-based (predictive) maintenance of equipment. Based on DCIM software, DMaaS aggregates and normalizes monitored data via an on-premises gateway. The data is encrypted and sent to a provider's cloud, where it is anonymized, pooled and analyzed, often using machine learning technology. In addition to real-time monitoring and reporting, key goals of DMaaS are to predict and prevent datacenter infrastructure incidents and failures and to detect inefficiencies or capacity shortfalls. DMaaS also ties into remote and on-premises maintenance services, enabling a full-service business model for suppliers. DMaaS takes the datacenter world beyond DCIM, and beyond single-site, proprietary management. 451 Research expects DMaaS offerings to speedily evolve throughout 2019 with richer features and upgraded machine learning or even deep learning models.

Distributed Resiliency

As enterprise IT spreads its work and data across a diverse, hybrid infrastructure, the underlying systems and services will require more complex software and network engineering. Some of the applications in the system will be cloud-native, but others will not. The result is that enterprises may now enjoy more functionality and more availability at lower cost – but at the price of an erosion in accountability, predictability and control. When failures do occur, sometimes in a gradual, partial way, and at other times on a grand scale, stabilization and recovery can take hours or days. As new edge services and systems are deployed, we expect the resiliency of complex architectures to be a growing concern, with some operators and providers getting ahead with new orchestration, analysis and monitoring tools, and service providers under pressure to be more open and responsive.

Due to these complexities, most enterprises will not reduce physical datacenter redundancy in the short term, as they first need to re-architect and test their high-availability applications for distributed resiliency. On the edge, however, 451 Research expects next-generation distributed resiliency mechanisms to emerge in 2019 and 2020 as early pilot projects go public in tandem with initial 5G radio network rollouts followed by industrial IoT applications that analyze and control mission-critical processes.

Recommendations

- **Take time to reconsider planning, design and build procedures.** If PFM datacenter systems do not play a part in an operator's infrastructure plans, there is a chance it will miss out on some potentially major improvements. Faster, more dynamic delivery of electrical, mechanical or fully PFM turnkey capacity changes the calculus of datacenter planning and construction. Recent years have seen a rapid evolution of PFM supply with vendors offering more options and considerable production experience with various installations from microsites to large-scale datacenter campuses.

- **Have a DMaaS strategy.** The key components of value for DMaaS will be a combination of DCIM software, cloud delivery, big-data analytics and on-premises field services. Most suppliers will be able to provide just one or two of these capabilities and will need strong partners elsewhere. In addition, the suppliers with the most data will dominate – something to bear in mind when evaluating suppliers.

- **Software providers should develop more tools for hybrid infrastructures.** There are already hundreds of tools and suites for monitoring IT and cloud services, but enterprises still need better ways to monitor and control their diverse and distributed digital infrastructure. Unification of facilities and IT management is still largely lacking due to organizational barriers and inertia.

Winners

- **Operators that are willing to adapt quickly.** Fast delivery and commissioning, granular capacity, and predictable cost and performance – made possible by manufacturing, as opposed to construction approaches – free up datacenter managers from the constraints of traditional planning and design practices, and free businesses from committing to large capital outlays for unknown future requirements. PFM technology lets operators respond better to different scenarios to reduce wasted capital and operational costs, while also introducing generational improvements in cost, operational efficiency and configuration choices.

- **Operators that understand the value of software.** The long-term trend is toward greater use of remote services, including those based in the cloud, with ever more intelligent and connected devices and analytics being used to predict problems or resolve them rapidly. The other clear direction is greater data sharing and even integration of facilities and IT management to optimize the infrastructure. Those that make the necessary organizational changes and work with their internal and external customers closely will be able to find major pockets of savings.

Losers

- **Operators with entrenched habits and no long-term technology outlook.** It will take years before competitive pressure from PFM systems, advanced DMaaS and distributed resiliency fully exerts itself on the market. However, operators that are not prepared for such a future risk are putting themselves at a severe disadvantage in time to market, cost and the ability to meet evolving customer requirements for efficiency and transparency.

- **Those overly focused on the endgame.** At the other end of the spectrum are operators and vendors that try to jump prematurely ahead of the times. Too much prescription from either technology vendors or datacenter operators, right or wrong, does not go down well with most customers, most of which seek continuity and dependability, not aggressive disruption, in the datacenter.

Trend 4: Rising Chip Power Will Force Rack Density Higher, Boosting Interest in Liquid Cooling

IMPACT TO THE MARKET

Implication: *As all major chip suppliers push to produce higher-power chips to attain more performance and overall system efficiency, air cooling appears to be more inadequate by the day. Some forward-thinking operators are looking beyond air-cooled datacenter designs to improve infrastructure economics. Direct liquid cooling (DLC) will likely make its way into non-HPC datacenters – gradually at first, then spreading rapidly. First-movers will likely gain overall cost and performance advantages.*

DLC, also known as on-chip cooling, is a method of heat dissipation where the processor or other components are close to or fully immersed in a liquid (dielectric). Over the past five years there has been a steady increase in the number of suppliers offering DLC: All major server OEMs now have a product. The majority of the take-up of DLC to date has been in high-performance computing (HPC) facilities, but there are signs that a wider range of operators will be persuaded by the improved server performance and lower capital costs. If DLC becomes more widely adopted, suppliers and operators of conventional datacenter air-cooling facilities would face a potentially sudden shift in direction and competitiveness of some sites.

The argument for DLC is simple: Liquids are orders of magnitude better at heat transfer than air. Because there are no fans, DLC is less noisy and creates much less vibration. For these reasons, DLC has been around for many decades; the first DLC systems appeared in the 1960s with IBM's System/360 mainframes. In fact, it was the primary form of cooling for datacenter IT when liquid-cooled mainframes dominated the server market up to the 1990s, mostly running on bipolar semiconductor technology. Vendors and customers alike trusted liquid cooling, a tested and understood technique, with the most mission-critical applications.

This switch over to more power-efficient complementary metal oxide semiconductor (CMOS) technology established air cooling as the norm in datacenters. Liquid-cooled electronics in the datacenter were relegated to exotic uses such as mainframes and some high-performance computing clusters; this remains the prevalent view among datacenter operators. Massive investments in air cooling systems by vendors and operators have created major inertia in the industry that resists change.

Yet 451 Research believes wider adoption of DLC will likely start happening in the next few years due to some key factors at play. The biggest is the evolution of datacenter processors, including GPUs and other accelerators. Processors in recent years have become ever more opportunistic in their performance. When a modern chip is below its critical temperature and power limits but one or more of its cores is highly loaded, it will start to increase frequencies (and voltage levels in support) in steps to meet the need for performance. This will continue until the chip hits a thermal or electrical limit. In this modern clocking regime, unlike in the static world of the past, more cooling capacity translates to more performance – sometimes considerably more: Most high-performance multi-core chips tend to be thermally, not electronically, limited (propagation delay in their circuitry).

All this is happening at higher power levels than before due to the nature of silicon technology scaling: the rate at which transistor shrink outpaces their reduced power use. No matter how much designers optimize circuitry for electrical efficiency, in the long term there is no escaping the underlying trend that a square inch of silicon needs more power with every generation. If the IT industry wants to take full advantage of even larger scales of integration (and if history is any guide, it will to drive better IT infrastructure economics), allowances for much higher power consumption per chip must be made. Notably, Intel's latest-generation Xeon processors have abandoned the previous limit of 165 watts (thermal power) per chip and introduced a 205-watt class for some of its server chips. In 2019, Intel will introduce even more powerful chips in a push for performance.

Even though air systems (heat sinks and fans) can still cool 205 watts, it is clear that they are reaching their practical boundaries and will become unviable for much higher-powered parts. DLC will help IT systems attain higher levels of performance while also making the overall infrastructure more energy-efficient.

The other consideration is the effort to lighten the capital burden of facilities. Optimization of datacenter designs has reduced the cost of building datacenters from around $25-30m per megawatt 15-20 years ago, to $10-12m today (and sometimes lower). Major operators, however, have encountered the fast-diminishing returns from further optimization within the current framework and realize it will take a more dramatic change in underlying assumptions to go further – one of which is the choice of cooling. With DLC systems, almost all the existing cooling infrastructure can be eliminated (chillers, air handling units, airflow management, humidification control, etc.) and 100% free cooling mode will become a reality in all climates. DLC systems also mean there are fewer components that can fail, and due to its more distributed and close-coupled nature the chances of a facility-wide cooling failure may become negligible, eliminating the need for capacity redundancy. All the while, IT failure rates should also drop as a result of better thermal stability, low humidity impact and lower average operating temperatures.

Most operators outside the HPC realm have in the past considered DLC systems inadequate for their requirements – and for good reasons. Many DLC systems were not engineered with datacenter scale operations in mind (points of failure, maintenance requirements, cost at scale, adding hazards, etc.), while IT vendors only offered lukewarm support. Based on feedback and requests from operators, vendors have addressed or started addressing most of these issues and DLC systems will soon be ready for prime time.

Recommendations

- **Explore DLC options now.** There are various forms of DLC systems on offer from a growing number of suppliers, each of which will offer (sometimes radically) different characteristics. There are not just server makers and established datacenter vendors on the market, but DLC specialists as well.

- **Have a technology roadmap.** The shift for most will not happen immediately in a wholesale fashion. The ability to install various DLC cabinets in a predominantly air-cooled datacenter will be a key requirement to meet tenant needs.

Winners

- **Operators willing and able to add DLC cabinets.** This will be a differentiator in the short term, particularly in areas where datacenter tenants tend to be early adopters, such as Silicon Valley (and datacenter markets around it such as Phoenix, Las Vegas and Portland) and Northern Virginia.

- **Enterprises and cloud firms that are able to take advantage of high-power chips.** The newer chips are currently designed for particular types of computation (e.g., data analytics), and enterprises may well be able to differentiate in their industry by taking advantage of these capabilities.

Losers

- **Datacenter operators that do not keep an eye on the DLC trend**. It may be years before DLC is widespread, but datacenter providers that do not consider how to add DLC may find themselves with an out-of-date, less appealing facility.

Trend 5: Climate Change and Sustainability Will Become Hot Issues

Implication: *Scrutiny of datacenter operators' environmental footprint and their social responsibilities will only grow with their resource consumption. Customers of MTDCs will also want to ensure their providers meet environmental standards and work actively to reduce their carbon and water footprints. At the same time, high summer temperatures and water scarcity in some locations mean that questions around long-term operational sustainability may arise more frequently – engineering based on historical worst-case climatic conditions might not be enough in the future. For reasons of policy, public relations and commercial risk management, operators need to plan for and actively manage their response to changes in the environment.*

The environmental footprint of datacenters has become a hot topic since the advent of massive multi-megawatt datacenter campuses in the second half of the 2000s. The unprecedented growth of datacenter infrastructures and their resource consumption, pulled by seemingly insatiable appetite for more digital services, hasn't gone unnoticed by policy-makers or environmental pressure groups such as Greenpeace. According to a 2016 report by the Lawrence Berkeley National Laboratory (LBNL) commissioned by the US Department of Energy, datacenters already accounted for about 1.8% of all electricity consumption in the US in 2014.

Even though LBNL's forecast is that overall datacenter energy and water footprint will barely grow by 2020, this is owing to the premise that resource consumption growth by hyperscale and large MTDC datacenters will largely displace enterprise footprint while energy efficiency gains will continue unabated. This is where our view diverges from LBNL's forecast: Cloud and MTDC datacenters have been adding net new capacity faster than they are displacing enterprise infrastructure. Also, the dynamic in the global market is different from that of the US: Latin America, the Middle East, Africa and large parts of Asia-Pacific are still catching up when it comes to digital infrastructures. At the same time, the energy efficiency gains of server chips are slowing as semiconductor technology scaling encounters more and more difficulties, while efficiency gains from virtualization taper off. As a result, 451 Research projects that available IT power in datacenters will have increased about 50% from 2014 to 2020.

Large operators tend to be more efficient thanks to better internal skills and the scale to go after even marginal savings, but more public scrutiny seems inevitable. The discussion has moved from the power usage effectiveness (PUE) of a facility – since most operators have optimized power distribution, airflow management and cooling modes – to questions around electricity sourcing (credits, power-purchase agreements or own on-site power) and water. However, trying to reduce carbon and water footprint can lead to additional non-trivial questions. For example, what should a datacenter operator do if it evaporates lots of water for cooling in a location where water is scarce, but grid power generation consumes even more water? Making the mechanical refrigerators work harder to reduce the need for evaporation could easily end up costing more water elsewhere. Conversely, if the grid runs on highly renewable energy, should a datacenter operator reduce water consumption and use more energy, at its own expense, to lower its overall footprint?

451 Research expects the topic to return to the desks of consultant and engineers, who will need to make choices based on stricter environmental mandates. Counterintuitively, data supports the case for relaxing operating temperatures in the data hall while also using efficient evaporative coolers – failure rates in most climates would not be negatively affected because the datacenter would spend more hours at lower temperatures (below 20 degrees Celsius or 68 degrees Fahrenheit) than at elevated temperatures. In fact, data indicates that if anything, adopting the wide band of 15-32 degrees Celsius (ASHRAE allowable class A1) for inlet temperature (59-89.6 degrees Fahrenheit) would in most cases help lower IT component failures while also enabling near 100% chiller-free cooling in most locations.

Though it is unlikely that most operators and their tenants would adopt such an approach, it supports the case for somewhat relaxed settings and optimization for more hours of free cooling. 451 Research also expects DLC to become part of the toolset in response to environmental challenges at large operators. Lithium-ion chemistries and grid demand response schemes (where the datacenter either feeds energy to the grid or reduces its load by partially energizing IT from batteries) will also score environmental sustainability points by helping intermittent renewable energy generation to enter the grid in larger quantities.

Finally, the environment is changing, and datacenter operators need to be prepared for natural disasters such as fires, floods, hurricanes, tornadoes, blizzards, etc., that have previously been considered rare or unlikely (e.g., once-in-500-years risk levels). Even when the datacenter is in a relatively safe location during a weather event, the datacenter staff and their families may lose their homes or be unable to reach the facility. Automation may help with some of these issues – during Hurricane Michael in Florida, all datacenter staff were forced to evacuate from some facilities and the datacenters functioned using only automated systems and remote monitoring for several days – as will redundancy built into software. However, automation requires investment for most operators, and for most of their clients the datacenter's resiliency is still their main or only bulwark against outages.

Recommendations

- **Operators should prepare an internal environmental report, if they have not already.** Examining everything from efficiency to power source to water use/source for each datacenter and summarizing how facilities are efficient and renewable, or not, will help prepare you for customers and officials that increasingly ask those questions. It also can lead to improvements. There may be relatively small investments that will boost efficiency, in some cases partly paid for by local power companies. Specialist firms are available that can help identify some of those options. Purchasing renewable energy credits may be possible as well.

- **Reexamine datacenter disaster plans.** Disaster planning must include worst-case scenarios for the datacenter, but also for staff and their families. Firms should prepare for broader geographic disruptions beyond those at the immediate datacenter location, and a wider range of potential disasters at or near the datacenter.

Winners

- **Operators that have already made efforts to boost efficiency and use renewables.** Datacenter providers that are already purchasing renewable energy credits and can point to specific areas of efficiency will appeal to the increasing number of customers focusing on environmental impacts. This will be a differentiator in the short run and may also help in discussions with local, state and national officials, particularly as an operator looks to expand and needs to convince officials that it is doing as much as possible to reduce environmental impacts.

Losers

- **Operators that are comfortable with their current disaster plans.** If you have prepared for natural disasters in your area, reexamine your plans with even more extreme disasters in mind – and not just weather-related. Earthquakes, riots, rising sea levels, smoke in the air – things you haven't thought of may strike. Firms will need to brainstorm with their teams on what could go wrong, how they would react and what processes and checklists could be put in place to help. Even if disaster does not strike, customers will increasingly see preparedness as a differentiator.

The Long View

Enterprises will continue to adopt public cloud (IaaS), but as one option among many. For at least the next several years, they will also want private hosted cloud, colocation and some on-premises equipment, and will make use of services for advice, migration and help managing various elements. In the longer term, the rise of IoT and massive amounts of data to be created, processed and stored will continue to push demand for datacenter space as well as for services. In this rapidly changing IT world, providers with the scale to influence customers and regulators, spread risk, reduce costs, offer more flexibility and automate will have a competitive advantage, as long as they know who their target customers are and still offer the service levels that customers want. There will be more M&A activity but also opportunity across many more geographies, so this industry should remain healthy for several years to come. We also expect more partnerships in the future will focus on providing additional reach for edge services, expanding the capabilities of more local providers.

There is a widespread assumption across the datacenter world that strong demand for IT, datacenter capacity and datacenter equipment will continue through the decade from 2015 to 2025. This, we believe, is broadly true and therefore the outlook is generally good for suppliers, assuming they maintain competitive positions. The patterns of demand for technology, almost all of which are upwards, will be shaped by many factors. Among these are supplier pricing, continued adoption of cloud, effectiveness of management and automation software (including for distributed infrastructure resiliency), the spread of the IoT, and macroeconomic and political factors, such as tax and privacy laws and energy policy.

Certainly, the landscape will not be even. Certain sectors of the market – such as hyperscale cloud providers – will grow more quickly, for example, but these operators will not necessarily invest as much in physical infrastructure (relative to the IT load) as traditional and smaller operators do today. Meanwhile, the latter group, along with enterprises, will fragment, with some growing, investing heavily and consolidating their specialist positions, while others will struggle to survive. It is likely that a greater proportion of technology sales will go to software, repeatable IT-based services and automation, and less (proportionately) to core power and cooling.

Several new and emerging technologies could also disrupt the economics and infrastructure of the modern datacenter – and the ecosystem of current suppliers. These include DMaaS, MMDCs, distributed resiliency, open-source infrastructure and datacenter microgrids, as well as others such as chiller-free cooling, software-defined power, DLC, storage-class memory and silicon photonics.

For the foreseeable future, the growth of public cloud and other third-party datacenter services will continue to shape the industry. While outsourcing offsets the demand for on-premises capacity, there will be minimal impact on the broader datacenter installed base. Looking forward to 2018 and beyond, there is good cause for executives and investors in datacenter technologies to be optimistic. While growth may not be even across suppliers and datacenter sectors, demand for datacenter capacity and services will continue to grow steadily and globally.

Further Reading

The way to win against the cloud is to play a different game, August 2018

From colo to cloud: Climbing the services stack - Cloud, IaaS and PaaS, March 2018

From colo to cloud: Climbing the services stack - Hosting, March 2018

From colo to cloud: Climbing the services stack - Storage and backup, March 2018

From colo to cloud: Climbing the services stack - Security, March 2018

From colo to cloud: Climbing the services stack - Managed OS, February 2018

From colo to cloud: Climbing the services stack - Introduction, February 2018

Suppliers of micro-modular datacenters think big by going small, February 2018

ABB, Rittal and HPE team to build micro-modular datacenters for industrial users, January 2018

Five ways to look at the top five datacenter markets in the US, October 2018

Prefabricated modular datacenters: from hype to delivery, August 2018

Price of cooling heats up as datacenter operators grapple with refrigerant shortages, June 2018

451 4SIGHT: Empowering the Digital Revolution, September 2018

Voice of the Enterprise: Cloud, Hosting & Managed Services, Workloads and Key Projects 2018

Voice of the Enterprise: Internet of Things, Workloads and Key Projects 2018

Index of Companies

About 451 Research

451 Research is a leading information technology research and advisory company focusing on technology innovation and market disruption. More than 100 analysts and consultants provide essential insight to more than 1,000 client organizations globally through a combination of syndicated research and data, advisory and go-to-market services, and live events. Founded in 2000 and headquartered in New York, 451 Research is a division of The 451 Group.

NEW YORK
1411 Broadway
Suite 3200
New York, NY 10018
P 212-505-3030
F 212-505-2630

SAN FRANCISCO
140 Geary Street
9th Floor
San Francisco, CA 94108
P 415-989-1555
F 415-989-1558

LONDON
Paxton House
Ground floor
30 Artillery Lane
London, E1 7LS, UK
P +44 (0) 203.929.5700
F +44 (0) 207.657.4510

BOSTON
75-101 Federal Street
5th Floor
Boston, MA 02110
P 617-598-7200
F 617-428-7537