Life Is On | Schneider Electric

Solving Edge Computing Infrastructure Challenges

White Paper 277

Revision 0

By Patrick Donovan and Wendy Torell

Executive summary

Edge compute (distributed IT) installations have become increasingly business critical. Deploying and operating IT at the edge of the network, however, comes with unique challenges. Solving them requires a departure from the traditional means of selecting, configuring, assembling, operating, and maintaining these systems. This paper describes a new, emerging model that involves an integrated ecosystem of cooperative partners, vendors, and end users. This ecosystem and the integrated micro data center solution it produces, help mitigate the unique challenges of edge applications.

RATE THIS PAPER ★★★★★

Introduction

In a time where everything, everywhere is being "digitized", we continue to see technologies like IoT, machine learning, AI, robotics, virtual/augmented reality, and data/video analytics being deployed. Businesses need these technologies to maintain competitive advantages and to stay relevant in their industries. But with these technologies comes latency, bandwidth, autonomy, and regulatory/security requirements that a centralized data center architecture cannot support. As a result, an end-to-end hybrid architecture, consisting of (1) local edge data centers, (2) regional edge data centers, and (3) centralized / cloud data centers, has emerged. And this hybrid architecture, including the local edge, must be highly resilient, as discussed in White Paper 256, *Why Cloud Computing is Requiring us to Rethink Resiliency at the Edge*. A simplified view of this architecture is depicted in **Figure 1**.

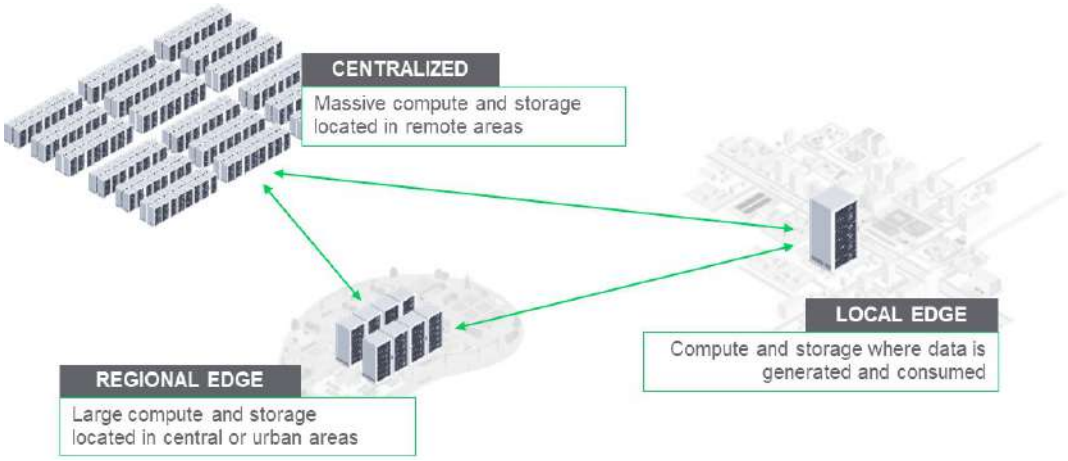

Figure 1
The hybrid data center architecture

That paper explains that although small local edge compute sites have traditionally not been treated the same as a central data center in terms of criticality and electrical redundancy, their availability is increasingly necessary for business continuity. As part of the digital transformation, new IT is being deployed on-premise to enable business critical processes and experiences that rely on network connectivity to the cloud. When that access point is down, employee productivity is limited, and the site's operations grind to a halt (i.e. can't deliver on product or customer experience).

A local edge site could be a manufacturing floor, a distribution warehouse, a bank branch, a retail store, a healthcare facility, etc. There are numerous cases of the technologies mentioned above being deployed across different verticals. Think about the technology of video analytics, for example.

In a retail store environment alone, we've identified many use-cases – from observing facial expressions of customers, to determining when queue lines will form, or even to eliminating the traditional "check out" process. In a manufacturing facility such as an automotive plant, it's being deployed for personal safety, for detecting production defects, and to streamline operations; In financial branches, to people count and identify demographics; and in a healthcare facility, to monitor patients and analyze when you need intervention. As technologies like this become a more integral part of the day-to-day business and/or customer experience, the edge compute sites that house the associated distributed IT equipment must be robust. The role of IT is no longer viewed as a cost center, rather it is tightly connected to the business strategy and to profit, making resiliency even more imperative.

There are two unique attributes of local edge environments, in contrast with regional edge or centralized data centers, that make it challenging to achieve the necessary

resiliency: (1) lack of IT and/or facilities staff on-site, and (2) a large number of sites, geographically dispersed.

In this paper, we focus on common problems found in edge environments, and what it takes to design, assemble, deploy, and manage a robust infrastructure, **so that business-critical applications remain available and personnel at the edge sites can focus on their primary tasks**. Specifically, we believe the answer lies in having an improved ecosystem of partners and tools throughout the lifecycle, and a fully integrated micro data center solution that is born from that ecosystem.

- **The ecosystem of partners** – Consider an organization with dozens or hundreds or even thousands of local edge sites (sometimes multiple sites within a single facility). How do I deploy the IT and support infrastructure in a standardized way, monitor the equipment, maintain the equipment across all those sites? Vendors, integrators, service providers, and the end user collaborating throughout all phases of the edge data center ensures the right equipment is deployed, the right management tools exist, and the right operations/maintenance protocols are established.

- **The Integrated micro data center** – This is the complete system that emerges from the ecosystem of partners. It includes the IT hardware/software (e.g. converged infrastructure), the physical infrastructure hardware, and software tools for monitoring/management of the edge environment. An effective solution is easy to deploy, easy to manage, ensures the IT is safe from physical threats, and remains resilient in the face of power disturbances and outages throughout its lifecycle.

For most companies with edge data centers, it is costly, inefficient, impractical, and unreliable to design, implement, and manage these sites without this integrated approach. According to World Wide Technology[1], a global IT solution provider, the ability to pre-configure technology platforms and devices before shipment to site:

- reduces field engineering costs by 25 to 40 percent
- increases order processing speed by 20 percent
- reduces maintenance costs by 7 percent

Challenges of edge sites

There have always been challenges related to selecting, configuring, deploying, and maintaining IT infrastructure at a given location. Those responsible for edge compute sites, however, often face the additional challenge of having to do this over multiple locations with little to no onsite staff. Working with the right ecosystem eliminates or mitigates the challenges described below.

Selecting and configuring infrastructure components can be complex. One mistake can propagate to hundreds of installations if not identified early on. Infrastructure parts must be selected that not only support the intended application, but that are also compatible with each other and fit each local site's conditions. This complexity leads to common mistakes we have seen at the edge such as:

[1] https://www.wwt.com/wp-content/uploads/2015/03/WWT_Integration_Centers_Overview.pdf, Accessed on February 28, 2019

- UPS undersized or grossly oversized for the expected IT load
- not considering UPS redundancy or UPS bypass (allows UPS replacement without downtime)
- not enough output receptacles (or wrong type) on rack PDU / UPS
- not enough U space in rack or spare space for expansion
- forgetting to order accessory parts kits needed to rack and cable everything
- not considering space constraints in rolling equipment into place

Deploying all the parts to each site can be a major logistical and workforce challenge. Some of the common problems we have seen include:

- finding wall space to mount cabinets onto studs
- sites not being prepared for large amount of de-trashing required
- separate components not arriving on time for the install
- losing small parts during onsite assembly
- difficulty obtaining necessary number of IP addresses
- site electrical not ready for installation of the IT equipment

Operating and maintaining multiple micro data centers from afar also presents unique challenges. Common concerns include:

- troubleshooting problems without IT or facilities staff on site
- comprehending large volumes of alarms and status change notifications
- not knowing there's an issue until something bad happens
- having to deal with multiple vendors when problems arise
- performing system upgrades and security patch updates
- knowing in real-time who is accessing the system locally and remotely
- standardizing maintenance service across multiple locations

Fundamentally, it is the unique challenge of having many sites and too few staff, combined with increasing criticality, that makes traditional edge IT deployments and operations untenable. An improved model is emerging that addresses these problems and challenges. This paradigm is based on embracing standardization, integration, cooperative partnerships, and cloud-based management tools. Integrated micro data center solutions emerge from this that include pre-configured IT, racks, power, cooling, security (physical & cyber), and management. Effective micro data centers are born from an integrated ecosystem of partners who work together to mitigate edge challenges found in each phase of the life cycle. The following sections explain what to look for in both the ecosystem and the micro data center solution that results from that ecosystem.

An integrated ecosystem of partners

To address the many challenges listed above, and reduce complexity, an integrated ecosystem of partners is crucial. **This cooperative network of people and tools is a new and emerging model that does not exist everywhere today. The burden is on the end user to seek out and partner with vendors, integrators, and service providers who embrace the principles and have the capabilities laid out in this paper.** This ecosystem should be centered around the end user and their technology needs for their business at the edge (**Figure 2**). Each has a specific set of functions within the ecosystem, and together they create not only the

complete integrated micro data center system, but also enable faster and simpler deployment, and more supported maintenance processes. This is particularly important for “lights-out” operations (i.e. no personnel onsite to support).

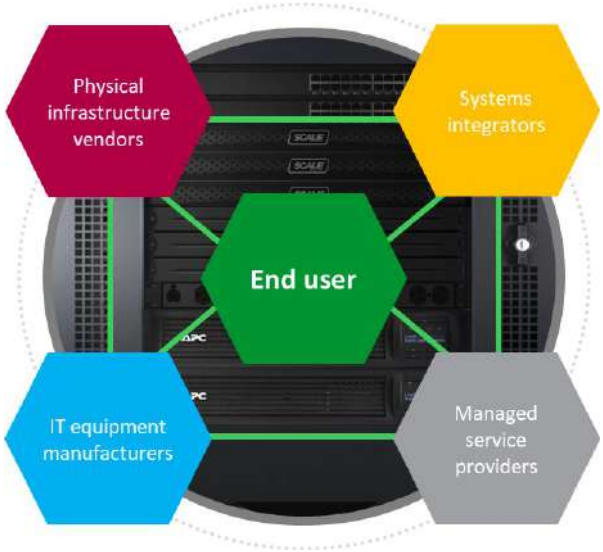

Figure 2
The integrated ecosystem of partners

Note, although the figure depicts five distinct entities, this is not always the case. Sometimes one partner fills multiple roles, or one set of roles is split up among two partners, or the end user fills some of the partner roles with in-house IT and services groups. But the functions and overall collaboration described in this paper remains.

Without the ecosystem of partners working together, it becomes very difficult to deliver and manage micro data centers across sites in a standardized, predictable, efficient, and reliable way. Together, they address configuration, integration, delivery, and operations/maintenance challenges.

End user: *Communicate business needs & challenges*

As the end user, your objective is around solving a business need – enabling an improved customer experience, increasing revenue, speeding up a manufacturing process, reducing head count at a distribution center, improving patient safety, and so on. You may have some specific plans to deploy technologies to meet those objectives, or you may rely on a consultant to identify the technologies and applications that can help you. For example, you may want to deploy video analytics in your retail stores to reduce queues, for example, but choosing the hardware and software to accomplish this should be handled by a partner, if there isn’t an inside resource or expertise. Many of today’s technologies are tightly integrated, which can add complexity. For instance, the video analytics may tie into an existing video surveillance system. Selecting and integrating the technologies takes an understanding of how these all fit together, and the right partner network can ensure this happens. It’s also important the end user shares information about each site’s conditions in terms of room use, access/egress situation, environmental conditions, availability of power and cooling resources (e.g., ducting), along with any other constraints.

IT vendor: *Drive interoperability through reference designs, tools, APIs, and certification programs*

The IT vendor(s) provide the servers, storage, networking gear, and software necessary to run the business applications. Converged, or hyperconverged infrastructure (HCI) is an integrated IT approach, that puts servers, storage, the network switch, and the firewall all in one hardware device. This can simplify the configuration and

testing since what used to take the expertise of multiple people is now done by one. Whether HCI or not, the IT vendors utilize their core competencies to configure IT hardware and software that is complementary with other vendors. The IT vendor should:

- provide a broad offering to meet a range of needs
- develop reference designs of IT based on specific edge use-cases or applications
- offer simple-to-use configurators to help tailor a solution for unique needs
- develop converged infrastructure solutions to reduce setup, install, and configuration burden on the integrator
- drive interoperability with other vendors hardware & software upstream and downstream of their hardware
- develop management tools with public APIs, so data can be accessible by any partner

Physical infrastructure vendor: *Simplify deployment & operations through configurators, reference designs, resilient infrastructure, and management tools*

The physical infrastructure vendor provides the equipment to ensure the IT remains secure and operational. This includes the physical enclosure (rack) for the IT equipment, the uninterruptible power supply (UPS) to provide battery backup for the IT equipment, rack power distribution (rack PDUs) to provide the needed outlets for the IT equipment, cooling, environmental monitoring such as temperature, humidity, & water sensors, physical security cameras & access control, and management software. The physical infrastructure vendor should:

- create rule-based configurator tools based on IT stack needs (see **Figure 3**)
- develop micro data center reference designs
- understand the criticality implications and risk of downtime
- provide offers that address appropriate level of environmental, physical security, and management
- develop management tools with public APIs, so data can be accessible by any partner
- Enable offsite integration and testing allowing products to be shipped configured without risking damage or warranty
- drive interoperability with other vendors hardware & software upstream and downstream of their hardware

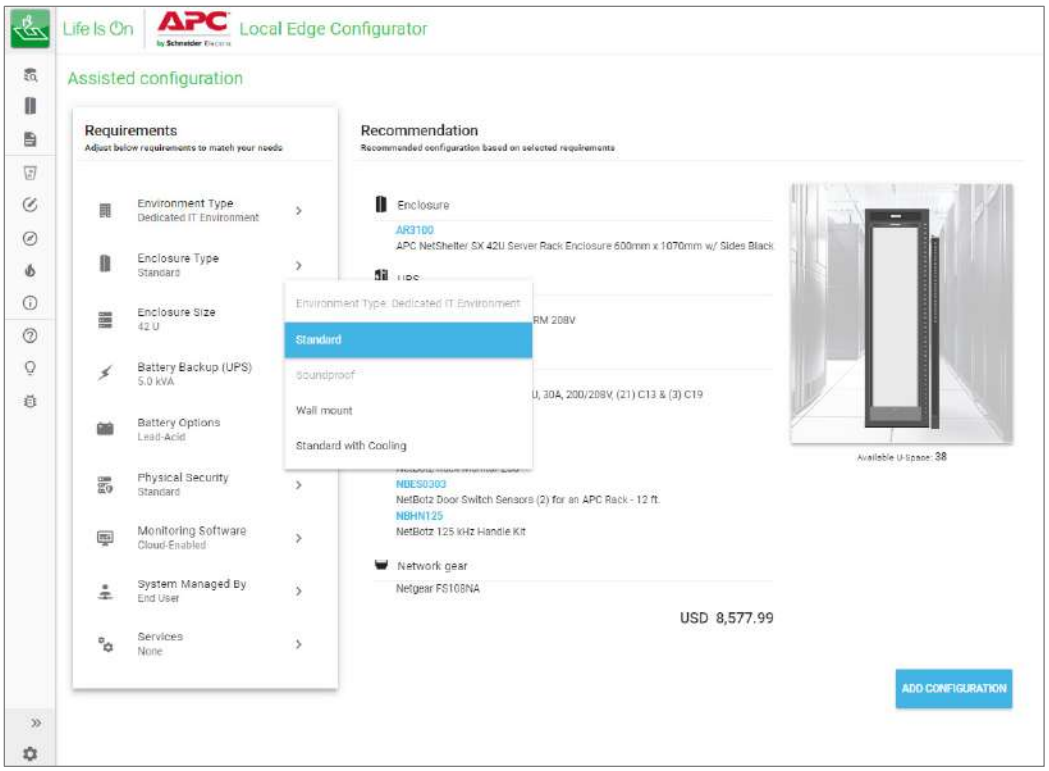

Figure 3
Example of a configurator for edge solutions

Systems integrator: *Add value through complete integration of IT hardware, software, and physical infrastructure*

The systems integrator is a company that brings together a range of solutions needed to meet the business objective. They serve a key role in the ecosystem, as they are the coordinator of all the parties involved. The systems integrator should:

- translate business need to IT requirements
- have expertise on the industry and technologies they deploy
- help determine the best IT stack needed to support the business application
- have a broad understanding of IT hardware and software integration
- have alliances with IT and physical infrastructure vendors
- help determine the physical infrastructure needs for the IT, based on power capacity, dimensions, weight, plug-types, criticality, etc.

There are lots of niche technology solutions that address specific business challenges or goals for the end user. The integrator will not be the expert of every vendor and every technology, rather have broad knowledge, and an ability to rely on partners for implementation when necessary. For example, a new robotics deployment in a warehouse designed to help improve efficiency of picking products off shelves must be integrated into the company's inventory management system. Tying into back-end data systems can be a significant project. The integrator will know the IT but will likely not understand the robotics.

Managed service provider: *Operate & maintain edge infrastructure through management tools and digital services*

A managed service provider (MSP), focused on information technology (IT) and/or operations technology (OT), is a company that helps operate the sites, addressing the staffing challenge on-premise. This often includes remote monitoring, proactive

maintenance, repair and upgrade services. This is where collaboration becomes key. The IT, OT, and physical infrastructure vendors need to offer management solutions with open APIs that the MSP can use or tie into. The MSP becomes the eyes and ears of the micro data centers. They implement digital services offered by vendors. **Everyone in the ecosystem needs the ability to look at the exact same data at the exact same time.** One partner may be troubleshooting the problem, the managed service provider may have to go and fix it, so it's helpful for everyone to work from the same data set. Cloud-based management offered by vendors provides the visibility and connectivity across the ecosystem. This is discussed in more detail in the following section.

The product of the ecosystem

When the ecosystem works as it should, effective integrated micro data center solutions are produced (as shown in **Figure 4**). Being essentially a complete "data center-in-a-rack", the physical infrastructure and IT systems are pre-configured, tested and, ideally, installed into the rack enclosure before delivery to the site. This simplifies deployments by reducing errors, site work, and disruptions to on-going business operations. Effective solutions improve operations by empowering managed services partners to provide better remote management and service capabilities.

Figure 4
Example of a fully integrated micro data center

There are three key elements of an effective micro data center solution. Each is described below.

1. Pre-configured and tested

Pre-configured means that parts and components have been designed by the manufacturer and validated by the integrator to fit and work together. This reduces human error when trying to select the right individual parts and components. It saves time and removes the "guess work" associated with designing from scratch. A tested system typically means that it has been built before. Either the vendor or a Systems Integrator who builds the actual micro data center will typically do a startup and test of the system to make sure everything performs as it should.
Pre-configured and commissioned solutions reduce or prevent onsite problems, while saving time and rework. Our experience with prefabricated data center modules show this savings to be on the order of 50% of the design/engineering time, as documented in White Paper 163, *Benefits and Drawbacks of Prefabricated Modules for Data Centers*.

Vendors should offer rules-based configurator tools that allow partners and end users to save time designing micro data center solutions. Specifically, these tools allow users to select and configure micro data center solutions in ways that are highly customizable yet **ensure only compatible selections are made** (e.g. rack PDU compatibility with UPS). Another way is through published reference designs. These are well-defined solutions for a given IT stack (e.g. specific hyper converged infrastructure solution) that include a design summary, bill of material, rack elevation, electrical one lines and floor layouts. These tools and designs can simplify and speed up the planning and design phases of a project to deploy edge and distributed IT. Ideally, reference designs are built into the vendor's configurator tool to provide an effective starting point for your business use-case.

2. Includes necessary infrastructure to maintain IT resiliency

Micro data center solutions should include all the needed physical infrastructure to support, power, cool, secure, and monitor the IT equipment. These items tend to be highly standardized, which makes them easily adaptable based on needs and constraints. An effective rule-based configurator helps select the appropriate size, quantity, redundancy of each part to ensure compatibility as a system. The following is a list of the key subsystems:

Rack enclosure – houses and secures the IT and supporting infrastructure. Racks should be capable of shipping and rolling in to the site fully assembled with everything pre-installed in the rack. This requires heavy duty "shock packaging" so that sensitive IT gear travels safely.

Rack power distribution unit (rack PDU) – distributes and controls AC power to the individual IT equipment. If individual outlet management is not needed and the UPS has the correct outlet configuration and voltage, rack PDUs may not be necessary. Furthermore, in cases where the UPS is hardwired, rack PDUs may not be necessary. Note, some UPSs offer outlet control built in.

Uninterruptable power supply (UPS) – provides battery backup power, voltage regulation, and surge protection to ensure uninterrupted operation of the IT equipment, regardless of what happens to the utility power. Given the criticality of the IT, a UPS is always recommended even if there is a backup generator present. For very critical loads and processes, some users provide a bypass switch to replace a UPS without experiencing downtime. In in some cases, users provide two power paths to the IT equipment and/or use two UPSs to ensure redundancy in case of failure or maintenance.

Active or passive cooling units – ensures IT equipment does not shut down due to overheating. Passive systems include ventilation features of the rack enclosure itself, as well as fan systems that help remove exhaust heat out of the enclosure into a return plenum or duct system. Active cooling systems may be needed for denser environments. This can be accomplished by direct expansion computer room air conditioners (CRACs), additional comfort cooling "mini-split" systems, or even chilled water-based computer room air handlers (assuming there's a chiller to reject the heat). See White Paper 68, *Cooling Strategies for IT Wiring Closets and Small Rooms*. Most small single-rack installations only require some sort of passive system.

Security & environmental monitoring – includes cameras, locks, and sensors for monitoring and protecting the micro data center from environmental threats. The critical nature of the IT, combined with a lack of onsite IT/facilities staff, makes this

an important aspect of the solution. This is exacerbated by growing regulatory requirements like GDPR, HIPPA, and PCI that are driving tighter security requirements.

3. Uses open APIs & cloud-based software management tools

Software management is critical for managed service providers and operations teams. The tools are necessary for having visibility and control from afar. However, having multiple sites lacking trained on-site staff, dictates a departure from traditional license-based, on premise software that has been commonly used for managing IT and data center infrastructure systems. New software suites have emerged that offer open APIs and take advantage of cloud, IoT, data analytics, and artificial intelligence technologies. **These new tools are what connect members of the ecosystem together to the operations phase of the micro data center. These new capabilities along with the MSPs who employ them essentially augment staffing for the end user by providing remote visibility and proactive control over all micro data center assets.** These new tools help solve edge and distributed IT operational challenges in the following ways:

- Cloud-based software enables unlimited scalability and automatic maintenance. For example, security patches and bug fixes can be automatically pushed by ecosystem vendors and implemented without onsite staff in a "no-touch" fashion.
- Open APIs provided by vendors empowers managed service providers to be able to pull necessary system data into their remote monitoring & management (RMM) tools to give them a complete view of the IT and infrastructure stacks from one place.
- Digital services connect MSPs and infrastructure vendors for improved maintenance. For example, pro-active servicing and replacing components could be carried out even before the local site knows there's an issue.
- Analytics & AI technology yields actionable insight for MSPs compared to traditional tools. For example, AI could predict when a battery needs to be replaced much further in advance, allowing replacements to be carefully planned. This is extremely valuable to geographically-dispersed fleet management as jobs can be scheduled at lower costs compared to emergency situations.
- Management of alarms and notifications is essential to providing a root cause for the problem or to make clear which alarms or devices are most critical and require attention. For example, a retail store chain may have thousands of UPS units deployed across the country. The manager could see hundreds of thousands of data points and status change notices from that population. It would be easy for critical alarms to go unnoticed or for a storm of alarms to occur with no understanding of what is driving them. Effective software tools would focus the operations manager on only the UPS units that need attention first or on the one alarm that started a cascade of further alarms.

These attributes combined help you maintain visibility and control over all your assets at the edge, ensuring the resiliency demanded at these sites today. Low to "no-touch" IT becomes possible. For more information on the capabilities of today's remote monitoring tools, see White Paper 237, *Digital Remote Monitoring and How it Changes Data Center Operations and Maintenance*.

Conclusion

In today's hybrid data center architectures, the edge of the network requires the same high resiliency as centralized or cloud data centers. But they are very different in two main ways: (1) they generally lack trained, onsite staff; and (2) there are multiple, distributed sites. This leads to some unique challenges that require us to think holistically and collaboratively when deploying IT.

We believe a collaborative ecosystem of vendors & partners must be chosen. This serves to augment the end user staff throughout the lifecycle of the sites – from configuration, to assembly and delivery, through operations and maintenance – enabling cost-effective resiliency. The right ecosystem produces a fully integrated micro data center that includes the IT, physical infrastructure, and management tools. Three key considerations to look for when planning IT deployments at the edge:

- Choose vendors that offer rule-based online configurators and reference designs to simplify and accelerate the selection and configuration and ensure interoperability with 3rd party systems; and provide resilient solutions that enable management for a fleet of dispersed sites.
- Choose system integrators that can create fully assembled and tested micro data centers that can be delivered complete. This ensures resources are optimized and reduces the risk of errors that can lead to delay and downtime.
- Choose MSPs that can tie into vendor's cloud-based management tools and digital services, so they can have the necessary visibility and control over the many assets that exist across a fleet of many IT deployments.

About the authors

Patrick Donovan is a Senior Research Analyst for the Data Center Science Center at Schneider Electric. He has over 20 years of experience developing and supporting critical power and cooling systems for Schneider Electric's IT Business unit including several award-winning power protection, efficiency and availability solutions. An author of numerous white papers, industry articles, and technology assessments, Patrick's research on data center physical infrastructure technologies and markets offers guidance and advice on best practices for planning, designing, and operation of data center facilities.

Wendy Torell is a Senior Research Analyst at Schneider Electric's Data Center Science Center. In this role, she researches best practices in data center design and operation, publishes white papers & articles, and develops TradeOff Tools to help clients optimize the availability, efficiency, and cost of their data center environments. She also consults with clients on availability science approaches and design practices to help them meet their data center performance objectives. She received her bachelor's of Mechanical Engineering degree from Union College in Schenectady, NY and her MBA from University of Rhode Island. Wendy is an ASQ Certified Reliability Engineer.

RATE THIS PAPER

Resources

Why Cloud Computing is Requiring us to Rethink Resiliency at the Edge
White Paper 256

Cooling Strategies for IT Wiring Closets and Small Rooms
White Paper 68

Digital Remote Monitoring and How it Changes Data Center Operations and Maintenance
White Paper 237

Browse all white papers
whitepapers.apc.com

Browse all TradeOff Tools™
tools.apc.com

Contact us

For feedback and comments about the content of this white paper:

Data Center Science Center
dcsc@schneider-electric.com

If you are a customer and have questions specific to your data center project:

Contact your Schneider Electric representative at
www.apc.com/support/contact/index.cfm