EATON
Powering Business Worldwide

Thought Leadership White Paper

The incredible shrinking data center

Strategies for reducing your footprint, saving money and meeting your power needs

By Hervé Tardy
Vice President and General Manager
Distributed Power Quality-Americas
Eaton Corporation

Executive summary

With companies relying on IT more heavily than ever, data center capacity requirements are steadily rising. Unfortunately, so are the costs associated with data center building and operation. As a result, organizations are increasingly searching for ways to reduce the size of their data centers without compromising their ability to meet business requirements.

This white paper examines the roots of that trend and discusses a variety of strategies you can employ to lower your IT operating costs by getting more computing done in a smaller physical footprint.

Table of contents

Why data centers are shrinking

The drive to reduce data center footprints stems from a variety of factors, including the five that follow:

High construction costs: Building data centers is an increasingly expensive proposition, as illustrated in the sample figures below for a representative 20,000 ft^2 (1,900 m^2) facility, based on a cost model developed by The Uptime Institute Inc., a research organization serving enterprise data center owners and operators:

Functionality Level	Power Capacity in Watts/ft^2 (m^2)	Total Budget
Tier II	50 (538)	$15,400,000
Tier IV	100 (1,077)	$48,400,000

Figure 1: All-inclusive budget required to build two hypothetical 20,000 ft^2 (1,900 m^2) data centers (Source: The Uptime Institute Inc.)

The Uptime Institute's cost model assigns a flat $220 ft^2 ($2,400 m^2) value to computer room floor space. However, since there's high demand for commercial real estate offering ideal conditions for data center construction—such as affordable access to electrical power and low exposure to natural disasters—that cost is likely to rise over time.

Mounting energy costs: Global demand for electrical power is rising rapidly, leading to steady increases in energy prices see figure 2. Since data centers can be voracious consumers of electricity, while smaller facilities tend to require less power, many organizations view shrinking their data center as an effective means of lowering their operational expenses.

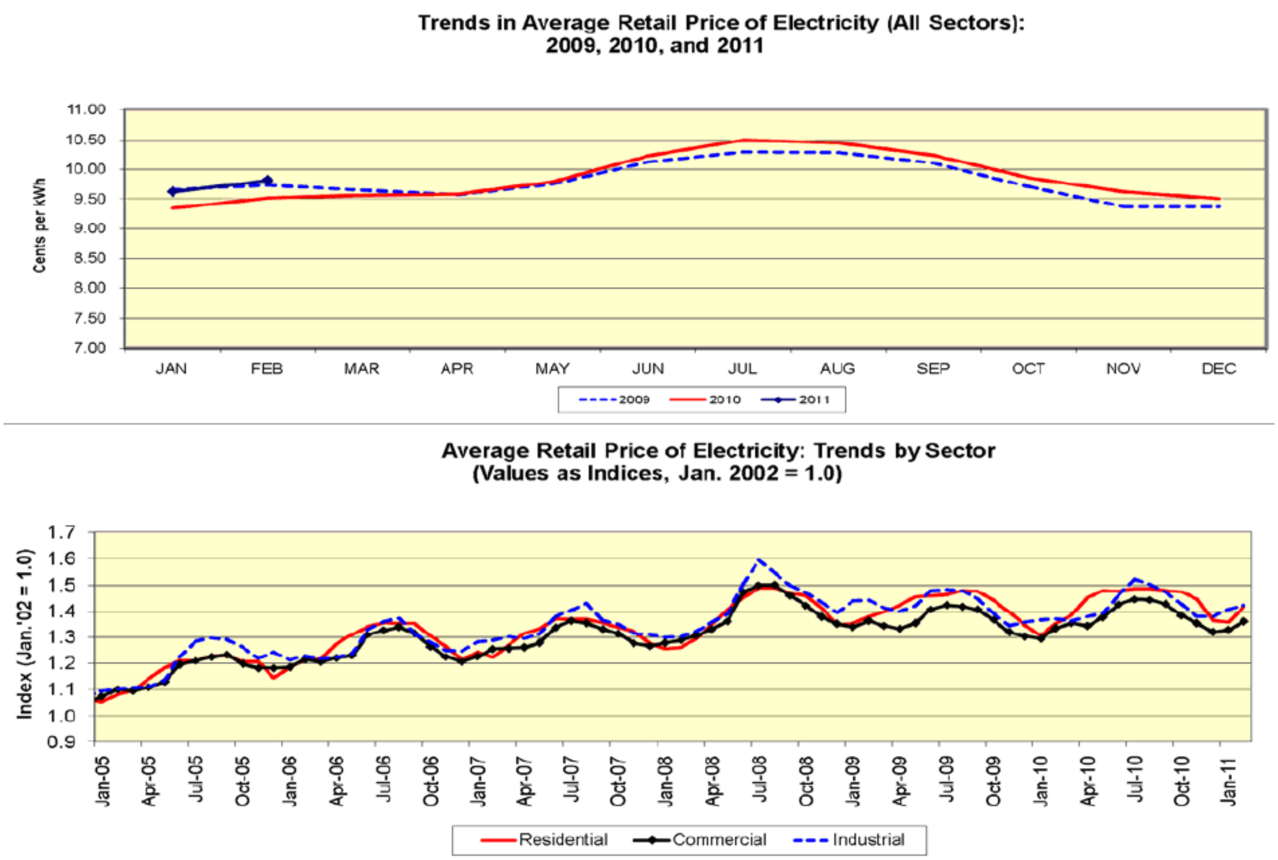

Figure 2: Average U.S. electrical power rates (Source: U.S. Department of Energy)

Steep water costs: Data centers consume enormous amounts of water. A 15-megawatt data center can use up to 360,000 gallons (1,362,748 liters) of water a day, according to data center designer James Hamilton, Amazon web services. Building smaller facilities helps organizations ease that expensive burden by reducing their water needs.

Intense budgetary pressures: Despite rebounding economic conditions, most businesses continue to maintain tight IT budgets. As a result, many IT and facilities managers are looking to lower equipment and operational outlays by building smaller computing facilities.

Shifting disaster recovery strategies: Looking to collect operational efficiencies, many companies are consolidating their data center operations into fewer, larger facilities. However, in an effort to reduce their vulnerability to power outages and natural disasters, other organizations are doing the exact opposite and maintaining multiple, smaller data centers rather than a single, large one. Still others are replacing in-house backup infrastructure with resources leased from co-location vendors and hosting providers. In the latter two cases, the end result is data centers that contain less equipment and thus require less space.

Strategies for conserving data center floor space

Companies that wish to join the trend toward smaller data centers can employ a wide range of techniques to shrink their computing footprint, including those discussed in the sections that follow.

Design-level strategies

Tailor your power architecture to mission requirements: A 2N power architecture features dual, identical power paths all the way from utility main to load. Typically, each path operates at not more than 50 percent of capacity, so if either one becomes unavailable due to a utility service interruption or required maintenance, the other one can keep the data center fully operational. Generally speaking, 2N power architectures are a wise choice for organizations with an extremely low tolerance for downtime, such as banks, utilities and emergency services providers.

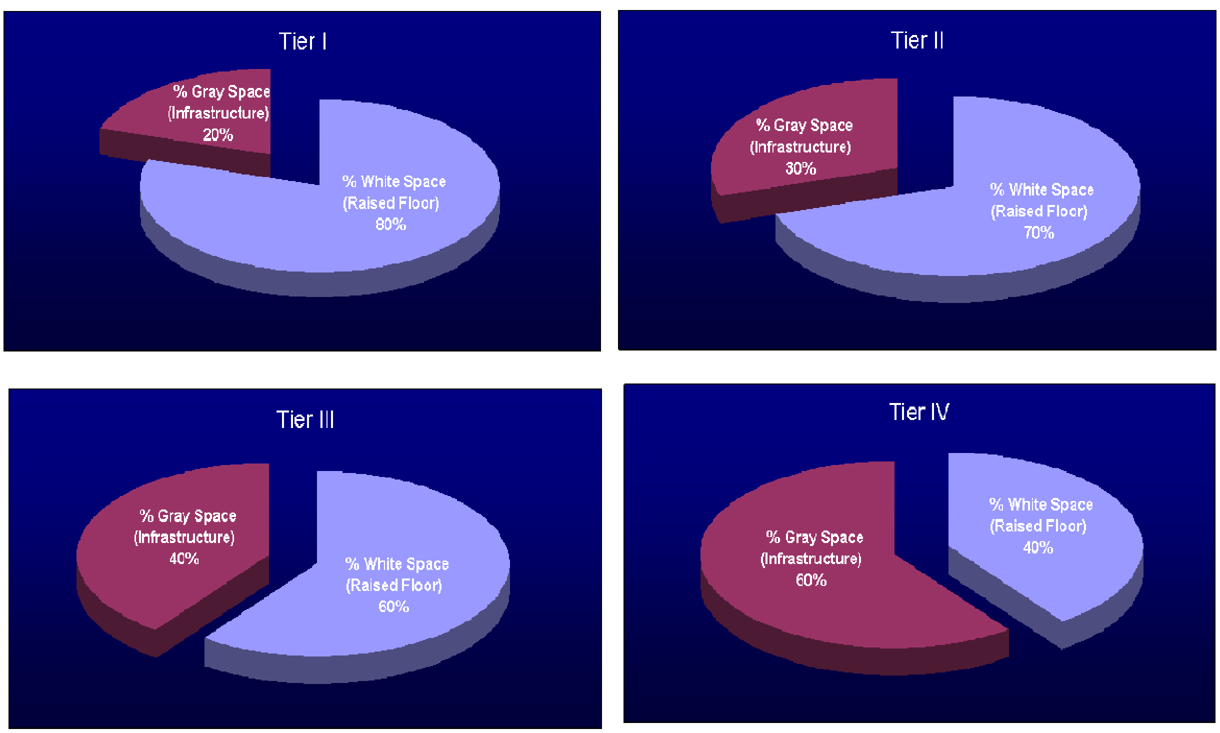

Figure 3: Space savings made possible by switching from a 2N power architecture to an N+1 architecture, by data center topology. (Source: The Uptime Institute)

As you would expect, deploying two of everything takes up a great deal of floor space. Companies with more mainstream uptime requirements can reduce their data center footprint while still keeping availability levels high by using an N+1 power architecture instead of a 2N configuration. An N+1 architecture includes one more of each power quality device than the minimum required to keep the facility up and running. That way, if any one device experiences an outage or needs maintenance, the remaining systems can still protect servers from data loss. This is a good application for redundant UPSs as well as redundant rack power strips and remote power panels (RPPs). N+1 configurations require far less space than 2N architectures, yet still provide adequate redundancy for most purposes, according to The Uptime Institute.

Figure 4: An N+1 configuration like this requires far less space than a 2N architecture, yet still provide adequate redundancy for most purposes.

Looking to compact their footprint even further, large data center operators such as Amazon and Google are increasingly using N and even N-1 power architectures. An N architecture doesn't include redundant uninterruptible power system (UPS) hardware, leaving critical workloads potentially unprotected in the event of UPS failure and during maintenance procedures. However, organizations that employ N architectures typically utilize state-of-the-art data mirroring techniques to synchronize multiple data centers on separate power grids. Should downtime occur at one facility, another one can keep mission-critical systems available to users. N-1 architectures take the same logic a step further by employing no UPS hardware whatsoever. Companies that use such configurations count on the availability of redundant facilities to compensate for the significantly higher likelihood of any one facility suffering data loss.

Though they dramatically lower data center floor space requirements, both N and N-1 power architectures are generally safe options only for the largest and most sophisticated data center operators. Furthermore, both architectures make sense only for companies with enough funds to build and maintain multiple data centers, even if one facility is all they really need.

Utilize 400V power: In the U.S., utilities typically deliver power at 480V. Most U.S. data centers, however, operate at 120V/208V. As a result, they must use a series of transformers to “step down” power from the 480V at which it's received to the 120V at which it's consumed by servers and other infrastructure devices. Operating a data center at 400V, instead of 120V/208V, reduces the number of transformations that must occur along the power chain. That, in turn, can lower floor space requirements by up to 60 percent, by eliminating the conventional PDU transformers. Further space can then be saved by utilizing busways for power distribution or 400/230V power. Manufacturers now offer UL-listed 400V UPS, RPP, busway and rack power devices.

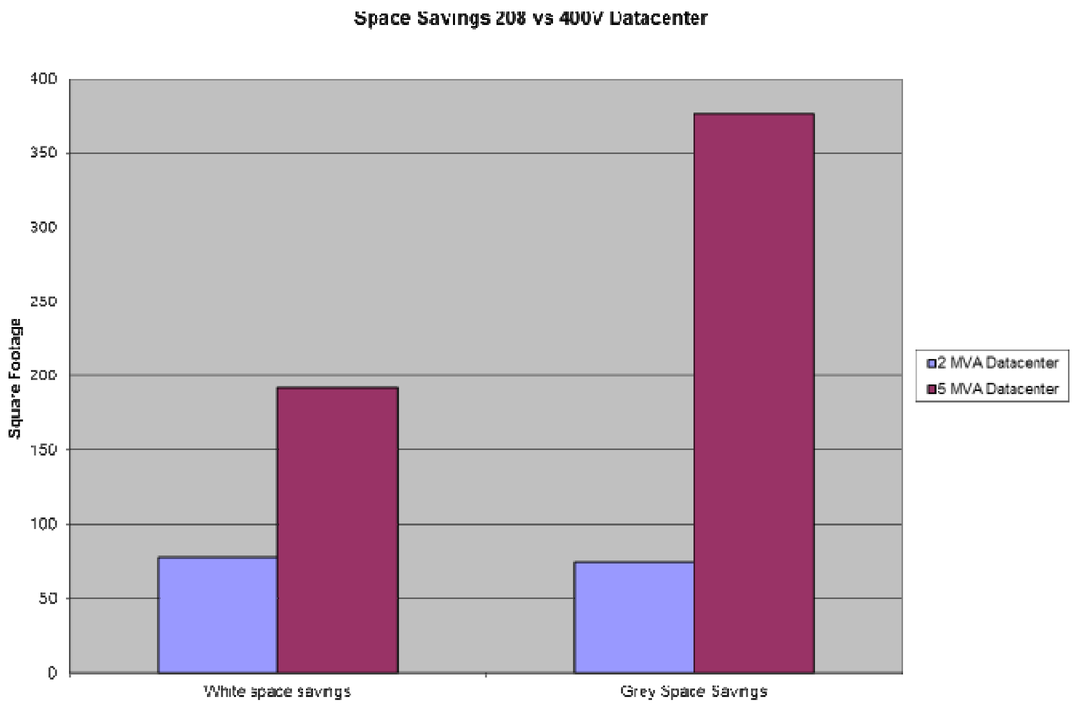

Figure 5: Switching to 400V power can substantially reduce data center floor space requirements.

As an added benefit, 400V data centers tend to be more energy-efficient as well, since fewer transformations generally translates into less power waste, and servers operate more efficiently at 240V as opposed to 120V.

Deploy containerized data center modules: Seeking to pack more power and computing capacity into less space, data center operators are increasingly utilizing containerized infrastructure modules. Some such products, from equipment manufacturers such as Oracle, HP, Dell and IBM, feature ready-to-use server racks. Others, such as the Eaton Uninterruptible Power Center, provide modular, pre-configured UPS systems. Yet another set of offerings include both the server hardware and associated power and cooling needed to run a miniature, self-contained computing facility. In addition to dramatically reducing the time required to expand or reconfigure an existing data center, all three module types enable companies to grow a data center up instead of out, shrinking its physical footprint by stacking modules on top of one another.

Figure 6: Stackable, containerized infrastructure modules like the Eaton Uninterruptible Power Center enable organizations to save data center floor space by building their facilities up instead of out.

Employ a “dark” data center design: To reduce IT operating costs, some companies are experimenting with dark, or unmanned, data centers. Such facilities rely on automation and remote management software in place of onsite technicians (see figure 7). Well suited to disaster recovery sites that get little regular use, dark data centers generally cost less to run than traditional ones. In addition, they also tend to be physically smaller, as they require significantly less office space and fewer parking spots than manned facilities.

White space strategies

Utilize virtualization: Virtualizing and consolidating servers enables data centers to conserve floor space by replacing large numbers of lightly utilized devices with a smaller number of heavily utilized devices. Consolidating storage resources through virtualization can similarly reduce a data center’s footprint by enabling it to house more data in less hardware. Leading-edge power companies can now integrate seamlessly with virtualization management systems like vCenter of VMware or SCVMM of Microsoft to enable live migration of virtual machines in case of power events. This is a significant improvement over the legacy server shutdown capability that is no longer addressing the need for business continuity which is inherent of any fully-virtualized environment. UPSs and advanced power management systems integrate seamlessly with software from VMware, Microsoft and other manufacturers to coordinate automatic migration or shutdown and infrastructure management in a virtual environment.

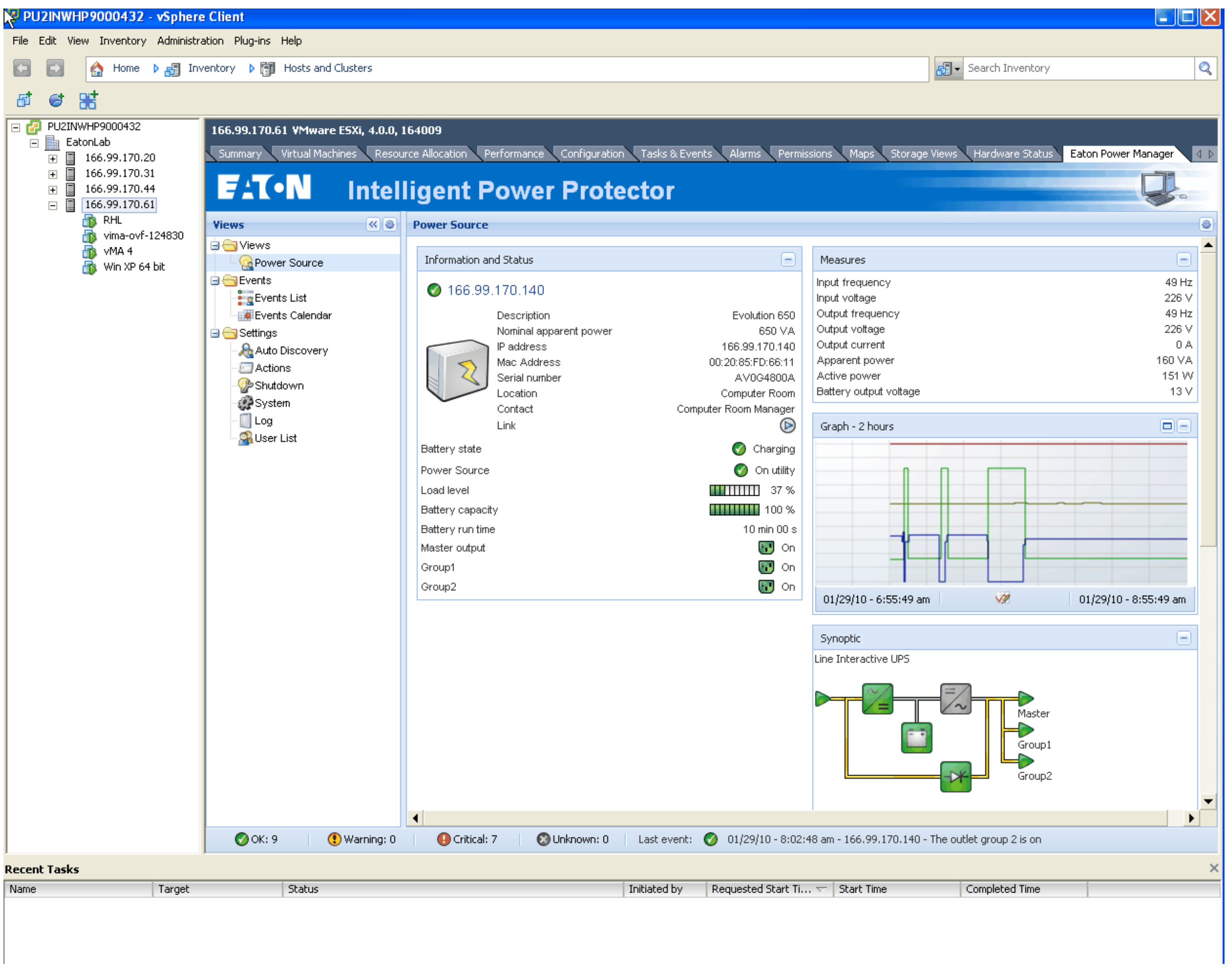

Figure 7: UPSs and advanced power management systems can be monitored through the VMware vCenter administration platform by simply opening a tab.

Deploy blade servers: Often used in conjunction with virtualization, blade servers are plug-and-play processing units with shared power feeds, power supplies, fans, cabling and storage. By compressing large amounts of computing capacity into small amounts of space, blade servers can dramatically reduce data center floor space requirements. They also enhance IT agility, since companies can simply plug in additional blades any time their processing needs grow.

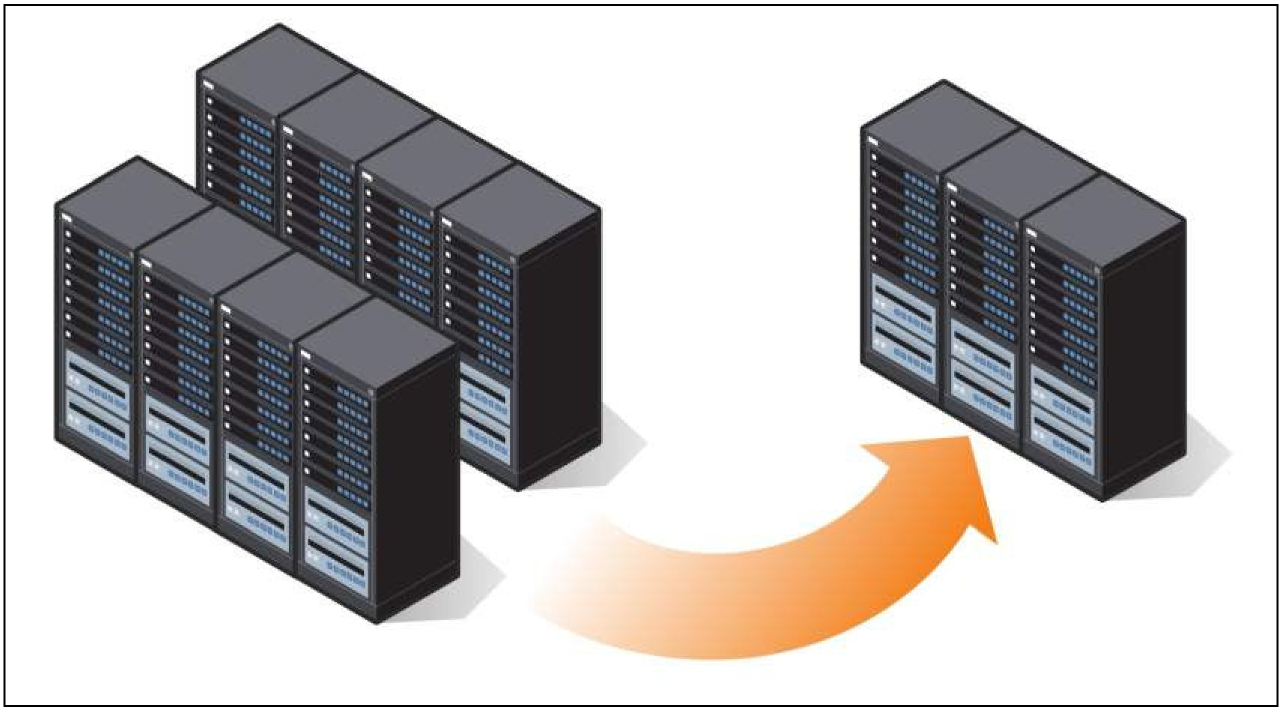

Figure 8: By virtualizing their servers, companies can get more work done on less hardware.

Leverage cloud computing: In an effort to lower overhead and increase IT efficiency, businesses around the world are rapidly adopting cloud computing solutions. Indeed, some 15 percent of all IT spending in 2011 will be tied to cloud services or infrastructure, according to analyst firm IDC.

Cloud computing can help organizations shrink their data centers in several ways. By offloading applications and infrastructure to externally hosted "public" cloud data centers that use the public Internet to exchange data, companies can reduce the number of servers they must own and maintain in their own facilities. Alternatively, businesses can deploy "private" cloud solutions that utilize the same basic technologies as the public cloud but reside on privately owned or leased servers. Through the sophisticated use of virtualization and automation, private clouds dramatically raise server utilization rates, thereby lowering floor space requirements.

Leverage capacity planning and asset management tools: Deploying more server and power resources than is necessary wastes floor space, but determining exactly how much capacity is required can be difficult, especially if virtual servers or a private cloud infrastructure are used. Capacity planning and asset management tools can help size a data center optimally for current and near-term needs, saving money while trimming the physical footprint.

Implement a passive cooling scheme: Today, most organizations dissipate data center heat by placing computer room air conditioning (CRAC) units around the periphery of their server floor. Unfortunately, CRAC-based cooling systems are often incapable of handling the greater power densities and temperatures associated with technologies such as virtualization and blade servers.

As a result, some companies are now investigating or implementing in-row liquid cooling systems, which take up floor space that could otherwise be dedicated to IT equipment. By utilizing a hot or cold aisle containment strategy instead of in-row cooling, data center managers can fit more server racks in their existing data halls while keeping operating temperatures at safe levels. One such strategy would be a passive cooling system, which can come in many different types; the most efficient enclosures are equipped with a sealed rear door and chimney that captures hot exhaust air from servers and vents it directly back into the return air ducts on CRAC units. The CRAC units then chill the exhaust air and re-circulate it. A properly designed passive cooling system can cost-effectively prevent even the densest, hottest server racks from overheating. Furthermore, some passive cooling schemes don't require raised floors, enabling companies to conserve the space formerly dedicated to maintaining air flow beneath their server enclosures.

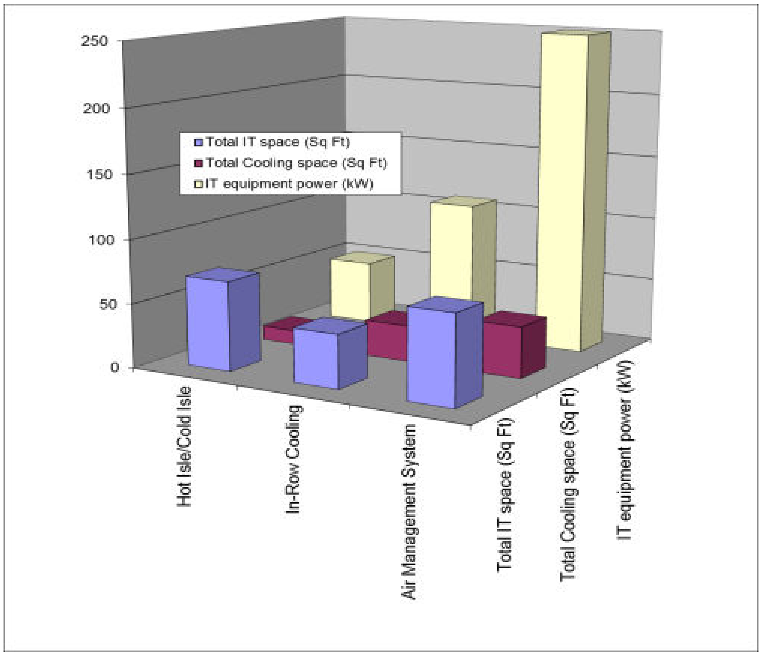

Figure 9: Passive cooling schemes enable companies to fit more computing power in existing floor space.

Figure 10: Passive cooling infrastructures use chimneys to contain and recirculate exhaust air

Switch from disk-based storage to static storage technologies: Though they currently account for a small fraction of storage hardware sales, static storage systems are slowly gaining popularity among enterprise IT managers. Such devices save data on solid-state memory chips, much like a USB memory stick. Though generally more expensive than conventional memory technologies, static storage systems are also dramatically faster, more energy-efficient and more reliable, since they contain no moving parts. In addition, static storage devices tend to be more compact than disk-based storage systems, so deploying them can help businesses save storage-related floor space in their data center. Tape-based archiving systems are also being impacted by cloud providers offering low-cost archiving storage at their facilities, something that can lead to reduction of equipment on the data center floor and still provide some type of disaster-proofing with data stored in another location.

Grey space strategies

Deploy modular, scalable UPSs and power distribution: Today, data centers often rely on large, inflexible UPS products that provide more capacity but operate less efficiently and take up more space than required. Using modular UPSs can eliminate that problem. Such products facilitate scaling a UPS deployment to meet current needs and expanding it incrementally as those needs increase.

For example, some modular UPSs provide up to 50 or 60 kW of capacity in 12 kW building blocks that fit in standard equipment racks. As requirements increase, another 12 kW unit can simply be plugged in. Even the largest UPS systems can be made modular in 200 to 300 kW increments. That's a scalable and efficient approach to keeping up with escalating power needs that lowers upfront capital spending and saves room in the data center.

Modular power distribution components offer similar economic and space-saving benefits. Sometimes called rack power modules, such products offer scalable "plug and power" distribution from a UPS or panel board to an enclosure-based PDU, eliminating wasted excess capacity and significantly reducing cabling requirements.

Employ alternative energy storage technologies: The lead- acid batteries most UPSs use are bulky. Employing alternative energy storage systems that require less space, such as flywheels, can help data centers reduce their physical footprint. A flywheel is a mechanical device typically built around a large rotating disk. During normal operation, electrical power spins the disk rapidly. When a power outage occurs, the disk continues to spin on its own, generating DC power that a UPS can use as an emergency energy source. As the UPS consumes that power, the disk gradually loses momentum, producing less and less energy until eventually it stops moving altogether.

Flywheels typically deliver only 30 seconds of standby energy, versus the five to 15 minutes of power generally provided by a lead-acid battery. However, research shows that more than 95 percent of utility outages last just a few seconds, so using a flywheel as a complement to batteries during brief power interruptions can save data center floor space by reducing the number of lead-acid batteries needed to protect server infrastructure.

Utilize air-side and water-side economization: Most data centers collect hot exhaust air and return water, chill it, and then re-circulate it. Facilities that utilize "air-side economization," however, simply pump hot internal air out of the building and pipe in cool external air. "Water-side economization" is a similar process in which return water is pumped through an external radiator or cooling tower rather than a chiller. Both techniques can significantly lower cooling infrastructure needs, reducing the amount of space required for CRAC units and other cooling resources. Moreover, studies have shown them to be viable options for at least part of the day even in mild or warm climates.

Raise server inlet temperatures: Data center operators typically keep internal temperatures at roughly 70°F (21°C). According to recent studies by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), however, most data centers can safely operate at temperatures as high as 80°F (27°C) and up to 60 percent relative humidity. Raising data center temperatures and humidity even a small amount can dramatically reduce the need for space-consuming cooling infrastructure.

Recommendations

Every business should examine its specific needs and constraints before deciding which space-saving strategies to employ. Broadly speaking, however, most organizations seeking to shrink the scale of their data centers will benefit from following these basic principles:

1. Make as much use of server and storage virtualization as possible, to fit more computing power and data into less space.

2. Look for additional ways to increase power and processing densities through the use of technologies such as blade servers, passive cooling and 400V power.

3. Follow modular design principles when building or retrofitting a data center to avoid implementing more infrastructure than is likely to be needed over the near or medium term. Look in particular for UPS, cooling and power distribution products that allow adding capacity incrementally as requirements grow.

Conclusion

Today's IT and facilities managers face a difficult bind: Though computing needs are constantly escalating, so is the cost per square foot of data center space and the price of critical supporting resources such as electricity and water. Consequently, more and more businesses are attempting to shrink the size of their data center without also shrinking its operating capacity.

Fortunately, there are many ways to compact a data center while still meeting business requirements. Most of them, moreover, are proven and cost-effective options for companies of almost any size. By studying the techniques described in this white paper and following the recommendations outlined above, you can get more done in less space and position your company to meet its IT requirements with maximum cost-effectiveness.

About Eaton

Eaton Corporation is a diversified power management company with 2010 sales of $13.7 billion. Celebrating its 100th anniversary in 2011, Eaton is a global technology leader in electrical components and systems for power quality, distribution and control; hydraulics components, systems and services for industrial and mobile equipment; aerospace fuel, hydraulics and pneumatic systems for commercial and military use; and truck and automotive drivetrain and powertrain systems for performance, fuel economy and safety. Eaton has approximately 70,000 employees and sells products to customers in more than 150 countries. For more information, visit www.eaton.com.

About the author

Hervé Tardy, 48, graduated from ESSEC Business School in Cergy-Pontoise, France and Stanford Executive Program. He is a 25-year veteran in the UPS industry and held multiple positions in sales, channel marketing, marketing communications, product marketing and product development. He started his career with the French UPS company Merlin Gerin, then Schneider Electric and MGE UPS Systems. His focus has always been to position the UPS as an IT peripheral more than a simple electrical device, and he turned out to become an expert in power management and software communication solutions. Hervé joined Eaton in November 2007 as Vice President and General Manager of their Distributed Power Solutions business unit, with responsibility over single-phase UPSs, software and connectivity products to reinforce the technology leadership of Eaton. His responsibility has recently been expanded to include the management of marketing and sales initiatives through the fast growing IT channel in the Americas. Tardy is based in Raleigh, North Carolina.