Powering Business Worldwide

Thought Leadership White Paper

Optimizing your Infrastructure for Cloud Computing

Best practices for managing a cloud IT environment

By Chris Loeffler
Data Center Applications Manager, Distributed Power Solutions
Eaton Corporation

Executive Summary

Cloud computing is generating enormous amounts of discussion and excitement in the world of corporate IT. Eager to drive efficiency up and costs down, organizations of every size and description are rapidly adopting Web-based software, platform and infrastructure solutions. Indeed, analyst firm International Data Corporation expects global spending by enterprises on cloud services to rise at a compound annual growth rate of 26 percent between 2009 and 2013, from $17.4 billion to $44.2 billion. Within a matter of years, experts predict, most businesses will utilize at least some cloud-based applications and services.

Yet for all its capacity to streamline management and boost agility, cloud computing poses unique power, cooling and availability challenges as well. To meet them, businesses must increase the strength and resiliency of both their electrical and mechanical infrastructures.

This white paper examines some of the forces behind rising adoption of cloud-based solutions, explores how cloud architectures impact data centers and discusses a series of concrete practices and technologies that can help companies collect the benefits of cloud computing without compromising uptime or overwhelming their power and cooling systems.

Table of Contents

Understanding cloud computing

Though definitions vary, cloud computing is basically the real-time delivery of IT infrastructure, services and software over the Internet or an internal Web-based infrastructure. Generally speaking, cloud solutions come in three broad categories:

- **Software-as-a-Service** (SaaS) solutions deliver software applications via the cloud. Salesforce.com, an online CRM application, and Google Apps, a Web-based productivity suite, are common examples of SaaS offerings.

- **Platform-as-a-service** (PaaS) solutions provide online access to computing resources and programming tools that developers can use to create and host cloud-based applications. Force.com, from Salesforce.com, and Microsoft Windows Azure are two familiar PaaS solutions.

- **Infrastructure-as-a-service** (IaaS) solutions offer infrastructure resources such as storage space and processing power over the Web. Amazon Web Services is among the best-known IaaS providers.

Regardless of which category they fit into, cloud solutions can be delivered via the "public cloud" or a "private cloud." Public cloud solutions exchange data over the Internet. They generally require no infrastructure on the user side beyond a Web browser and high-speed Internet connection, and are typically billed under a subscription-based, pay-as-you-go model in which you pay only for the resources you use. Private cloud solutions utilize the same basic technologies and protocols as their public counterparts, but reside on privately-owned or leased servers located behind a corporate firewall. Many IT managers believe that makes them easier to secure and manage.

Both public and private cloud solutions offer considerable business and technical benefits:

Efficiency: Cloud solutions generally run on flexible, highly automated infrastructures that are simpler—and hence less expensive—to manage. Additionally, since they don't require local server resources, public cloud solutions free businesses from heavy upfront investments in hardware and software, turning IT from a capital expense into an operational one.

Agility: Most cloud infrastructures make extensive use of server virtualization, a technology that allows a single physical server to host multiple "virtual servers," each with its own operating system and applications. Since those virtual servers are essentially little more than sophisticated computer files, creating a new one typically takes a fraction of the time required to setup a new physical device. As a result, cloud environments empower businesses to deploy new applications and services far more quickly than conventional ones.

Scalability: Most cloud solutions capitalize on the flexibility made possible by virtualization to let users add or remove processing and storage capacity dynamically in response to fluctuating needs.

How cloud computing impacts IT infrastructures

Maintaining availability and providing adequate power and cooling are long-standing challenges for managers of traditional data centers. However, for a variety of reasons, addressing those issues can be even harder for managers of cloud data centers.

Power and cooling

The server hardware that most cloud infrastructures use to host virtual machines is bigger and more robust than a typical single-function server. It's also far more heavily utilized: While the average non-virtualized server operates at perhaps 5 to 15 percent of processing capacity, the average virtualization host server may be as much as 80 percent utilized at any given time. For both reasons, the virtualization host servers in

most cloud data centers demand more power than conventional servers, and put greater strain on power distribution units (PDUs), panelboards and uninterruptible power systems (UPSs).

This is particularly true when organizations employ blade servers to host their virtual machines. Blade servers utilize multiple plug-and-play processing units that share common electrical feeds, power supplies, fans, cabling and storage. Such an arrangement conserves data center floor space, simplifies hardware management and enhances IT flexibility. However, it also significantly raises compute densities. As a result, blade servers generate enormous amounts of heat and radically increase rack-level power requirements. Indeed, while a typical rack of conventional servers might draw four to six kW of power, a typical rack full of blade chassis can draw as much as 30 kW. That's more than many power and cooling systems can handle.

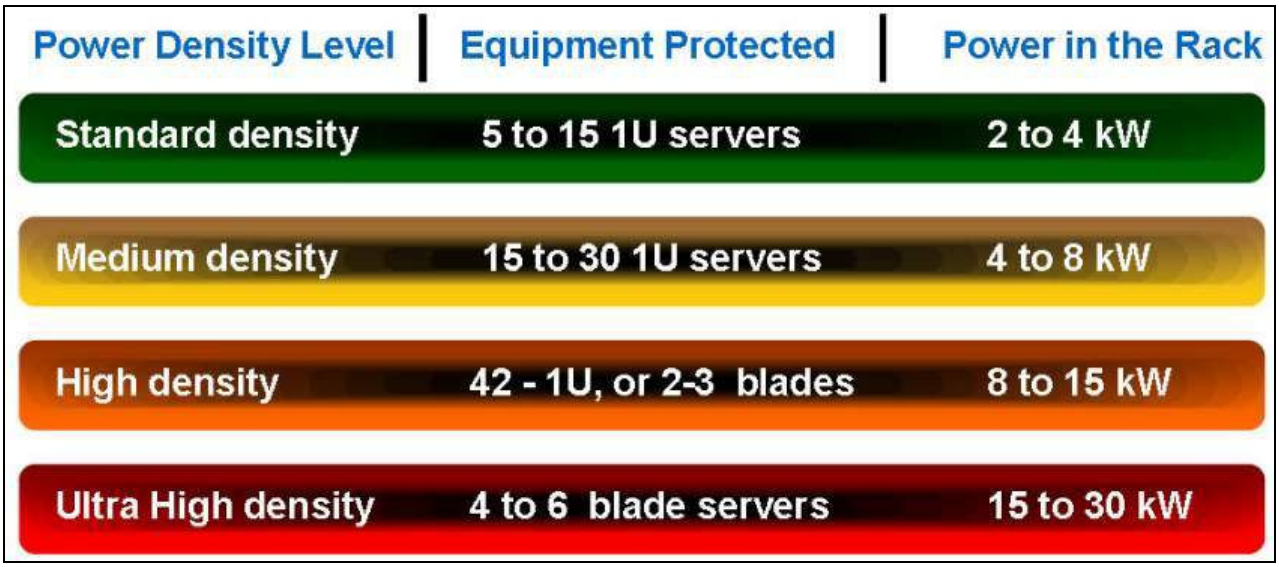

Figure 1: Most cloud infrastructures are ultra high density environments with significantly greater rack-level power requirements than conventional data centers.

Reliability

The heavy use of virtualization in most cloud computing environments also has implications for availability. In a traditional data center, each server typically supports one application. In a virtualized cloud computing data center, a given host server may support a dozen or more applications. Any time a host machine fails, therefore, it has the potential to impact large numbers of users and business functions.

Moreover, the flexibility that helps make cloud computing and virtualization so attractive can also inadvertently produce downtime. Without proper administration, shifting workloads suddenly within and across data centers can result in overloaded circuits or overtaxed cooling systems, which can in turn bring down critical systems.

Strategies for powering and cooling cloud-based infrastructures

IT and facilities managers can meet the significant challenges of providing power and cooling to cloud computing environments by implementing strategies like those discussed below.

Use modular power and cooling system components

No cloud data center manager ever wants to be caught with less power or cooling than they need to meet rising customer demand. On the other hand, deploying excess capacity far in advance of future requirements wastes time, money and resources.

Using modular power system components is a smarter approach. Such products let you add capacity quickly and incrementally as your needs increase. For example, a modular scalable UPS for a small cloud environment may provide up to 50 or 60 kW of capacity in 12 kW building blocks that fit in standard equipment racks. As your requirements increase, IT personnel can simply plug in another 12 kW unit, growing capacity (in this example) from as little as 12 kW up to 60 kW N+1. That's a scalable and efficient approach to keeping up with escalating power needs that's far more economical than purchasing surplus capacity in advance. Moreover, rack- based modular power system components tend to be compact and easy to install, making them an ideal fit for fast-paced cloud data centers, in which technicians are constantly moving, changing and adding infrastructure resources.

UPS power building blocks typically come in sizes as small as 3 kVA up to sizes in excess of 1,000 kVA. In general, organizations should employ UPS building blocks that are about four to six times smaller than the expected full capacity of the finished block.

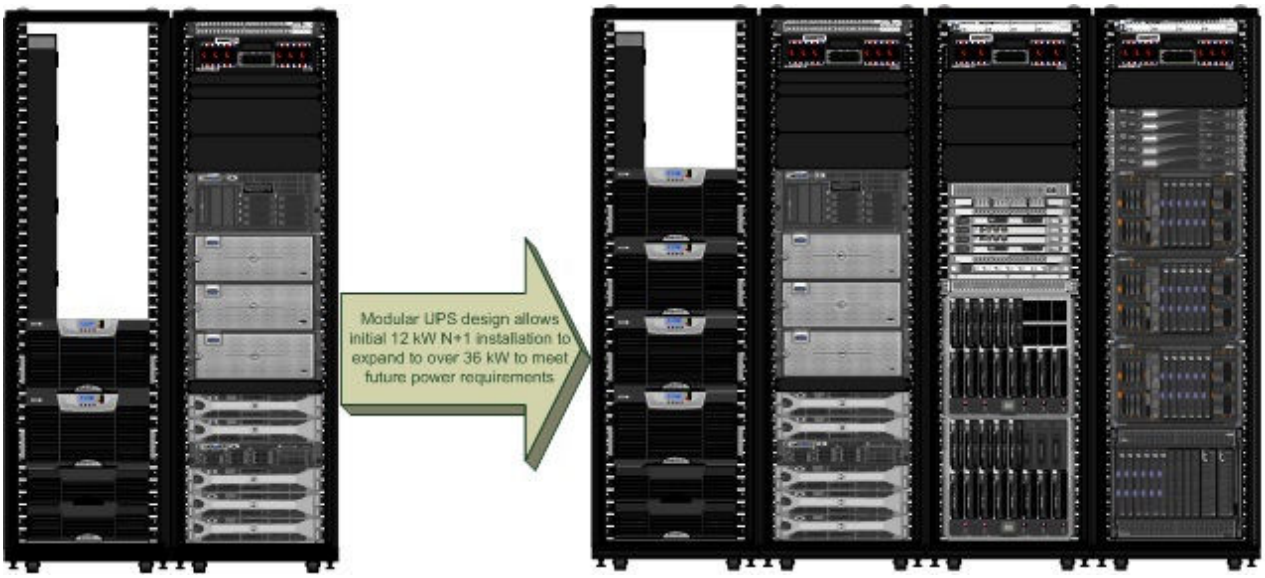

Figure 2: Modular power components let you add capacity incrementally in response to rising demand.

Deploy a passive cooling system

Today, most organizations dissipate data center heat by placing computer room air conditioning (CRAC) units around the periphery of their server floor. Many companies also use "hot aisle-cold aisle" hardware configurations, in which only hot air exhausts or cool air intakes face each other in a given row of server racks. That produces convection currents that generate a cooling, continuous air flow. However, while technologies such as these are usually more than sufficient for traditional data centers, they are often incapable of coping with the searing heat produced by cloud infrastructures. Thus, public and private cloud environments typically require newer and more robust cooling technologies.

Modular cooling system components, similar to the UPS designs discussed above, are currently available from a variety of manufacturers. However, deploying additional cooling blocks as needs increase is not as simple as with modular UPSs. Data center managers must typically install costly under-floor or overhead piping for those units in advance. As a result, many companies prefer to reduce cooling-related power consumption by equipping their CRAC and computer room air handler (CRAH) systems with variable frequency drives or electronically commutated (EC) fans. Variable frequency drives save energy by enabling air handling systems to run slower when servers require less cooling and faster when workloads are at their peak. Similarly, EC fans use "intelligent" motors to run faster or slower as needed based on airflow demand.

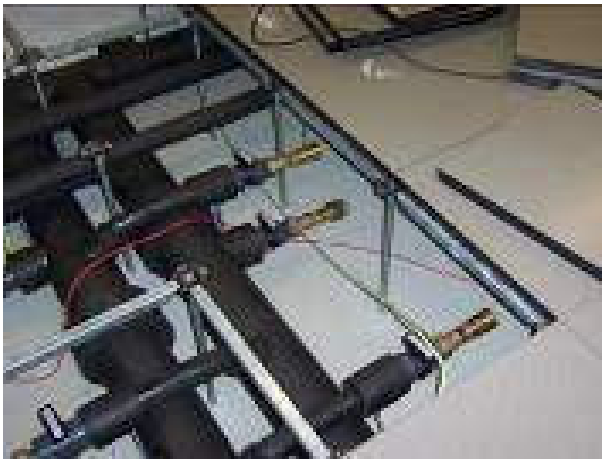

Figure 3: Under floor or overhead piping with quick connects can be added during initial construction, though they can increase construction costs.

Companies looking for even lower upfront costs and higher operating efficiencies can install passive cooling systems. These employ enclosures equipped with a sealed rear door and a chimney, which 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. Passive systems typically require a strong air flow “seal” from the front of the cabinet to the rear so that only minimal hot server exhaust air mixes with incoming cool air from the CRAC units. By segregating hot air from cool air more thoroughly than ordinary hot aisle-cold aisle techniques, a properly-designed passive cooling system can cost-effectively keep even a blazingly hot 30 kW server rack running at safe temperatures.

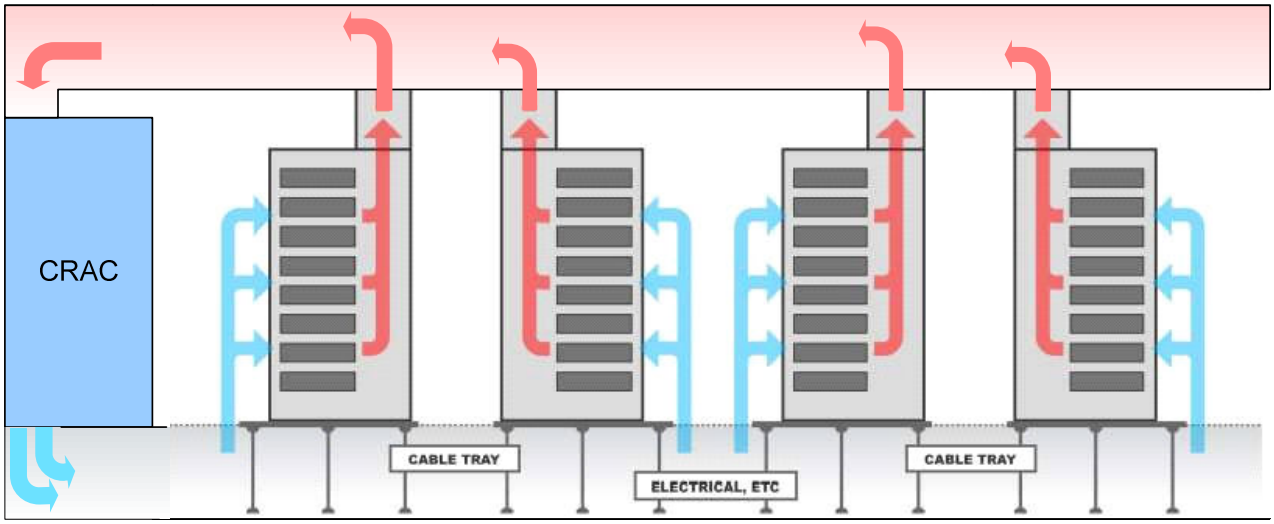

Figure 4: Passive cooling systems like the one depicted here utilize sealed enclosures and chimneys to vent hot air out of the data center before it can mingle with cool air, resulting in more efficient temperature management.

Construct multiple facility rooms

Large data centers like those that supply public cloud services often house UPS equipment in a dedicated facility room adjacent to the server floor. Setting up *two* facility rooms, one for UPS and power system electrical components and the other for UPS batteries, can be an even more efficient arrangement. While UPS electronics can typically operate safely at 35°C/95˚F, UPS batteries must usually be kept at 25˚C/77˚F.

Placing UPS batteries in their own environmentally-controlled room reduces the amount of heat your cooling systems must handle. Deploying higher heat-generating equipment in elevated, but acceptable, temperature environments helps lower your cooling needs and power bills.

Strategies for improving a cloud infrastructure's reliability

Though cloud computing can make preserving uptime more difficult, practices and technologies like those discussed below can significantly ease that task.

Conduct a power chain audit

Organizations planning to add a cloud infrastructure to an existing data center should include a thorough power chain audit in their pre-deployment planning. When conducted by a certified power system engineer, a power chain audit can help you evaluate your power systems and determine which, if any, should be upgraded, augmented or modernized to support a cloud environment's more demanding requirements. In addition, a power chain audit can also assist you in identifying potential cost-saving opportunities and assessing the electrical safety of your power components.

Add redundancy to your power architecture

Organizations can increase availability and reduce the likelihood of unscheduled downtime by utilizing a redundant power system architecture, such as these:

N+1: An N+1 architecture includes one more UPS, generator or other power component than the minimum required to keep server equipment up and running. Thus, if any one component experiences an outage or requires maintenance, the remaining systems can still provide adequate protection against data loss. An N+1 architecture is often sufficient for the needs of a small or medium cloud environment.

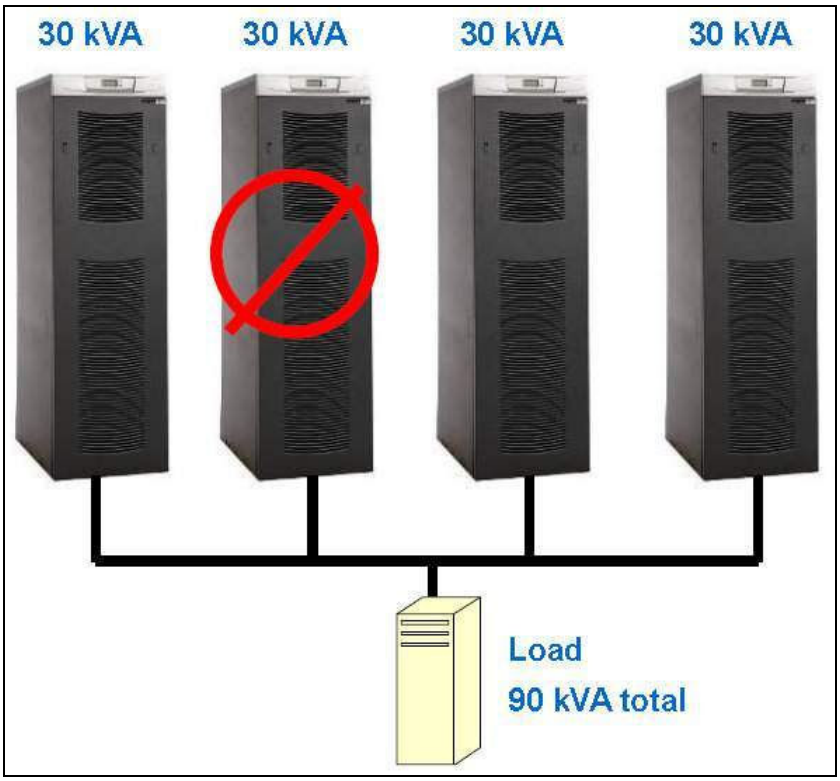

Figure 5: An N+1 power architecture continues functioning even if a UPS goes offline due to technical problems or maintenance requirements.

2(N): A good choice for large cloud environments, 2(N) architectures feature two separate but identical power paths, each of which is capable of supporting an entire infrastructure on its own. Under normal conditions, both paths operate at 50 percent of capacity. Should one path experience planned or unplanned downtime, however, the other can compensate by temporarily running at 100 percent of capacity.

2(N) architectures are often used in conjunction with servers that contain dual power supplies. In such an arrangement, each power supply typically utilizes a separate power path. That way the server remains available even if an entire power chain goes offline for repairs or maintenance.

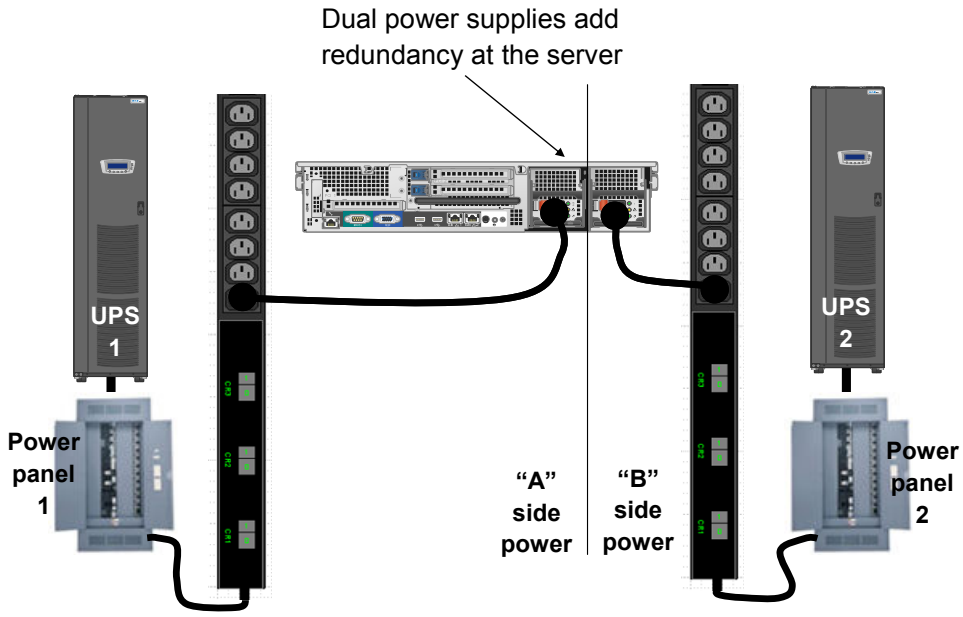

Figure 6: A 2(N) power architecture utilizes separate and identical power paths. If either path goes down, the other compensates automatically.

Deploy replication software

To further improve reliability, cloud data center managers can use software-based redundancy techniques such as replication. Replication solutions continuously capture changes as they occur on protected servers and then replicate them in near real time to backup servers. If a primary server fails, the backup server can swiftly step in to ensure seamless availability.

Utilize live migration software

Capitalizing on the live migration functionality built into many server virtualization solutions is another effective software-based reliability strategy. Live migration systems like VMware's vMotion solution enable administrators to move virtual servers almost instantaneously from one physical host to another in response to technical issues or maintenance requirements. If a physical server shows signs of impending failure, for example, technicians can use live migration software to swiftly transfer its virtual machines to another host machine until the problem is resolved.

Employ integrated management software

Clean, dependable power is as critical to the successful operation of cloud infrastructures as processing capacity and storage space. In order to keep a cloud data center running smoothly, then, administrators need complete, up-to-the-minute information about the status of both their IT resources and their power resources. Today, many cloud operators use separate management tools to monitor their server and power environments. However, integrated solutions are now available that allow administrators to manage physical servers, virtual servers, UPSs, PDUs and more all through a single console.

For example, the latest releases of Eaton's Intelligent Power Manager software and VMware's vCenter Server virtualization management system work together to provide a comprehensive view of network- and power-related events and alerts in one place. What's more, administrators can configure the two systems to act on status information dynamically. Should your data center experience an electrical outage, for example, Intelligent Power Manager and vCenter can gracefully shut down affected virtual and physical servers before your UPS systems run out of battery power. Alternatively, if your data center is equipped with vMotion, you can automatically migrate impacted virtual machines onto host devices in another, fully-operational facility. Either way, the ultimate result from the end user's perspective is continuous uptime.

Conclusion

Cloud-based software, platform and infrastructure solutions improve the efficiency and elasticity of IT operations. As a result, large numbers of businesses are tapping into public and private clouds today, and even greater numbers will join them in coming years.

However, cloud computing subjects data centers to significant new pressures. Cloud infrastructures make extensive use of virtualization and higher powered servers including blade servers, technologies that dramatically increase rack-level power and cooling requirements. Moreover, cloud data centers tend to be dynamic environments in which virtualized workloads migrate freely among physical hosts. That increases IT agility but can also result in blown circuits and other electrical problems that lead to service outages.

To master these challenges, organizations should adopt technologies and techniques that increase the reliability and redundancy of their physical and virtual environments, including power and cooling systems. These can range from modular power components and passive cooling schemes to data replication solutions and live migration software. In addition, proper monitoring and control of physical and virtual systems will help organizations manage their infrastructure more easily. Together, such tools and strategies can help any company enjoy the power of cloud computing reliably and cost-effectively.

About Eaton

Eaton Corporation is a diversified power management company with 2009 sales of $11.9 billion. 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

Chris Loeffler is the global applications manager for Eaton Corporation, specializing in data center power solutions and services. With more than 19 years of experience in the UPS industry, he has overseen product management of more than 20 UPS products for data center and industrial applications. Mr. Loeffler has held a variety of positions with Eaton, including roles in service engineering, application engineering, and more than 10 years within product management. Mr. Loeffler has authored a number of articles for trade publications and written several white papers on energy efficiency in the data center. He has also written articles on various UPS topologies for data center and industrial applications.

Tutorials on demand
Download Eaton white papers to learn more about technology topics or explain them to customers and contacts. Maintenance bypass, paralleling, UPS topologies, energy management and more are demystified in free white papers from our online library: **www.eaton.com/pq/whitepapers**.