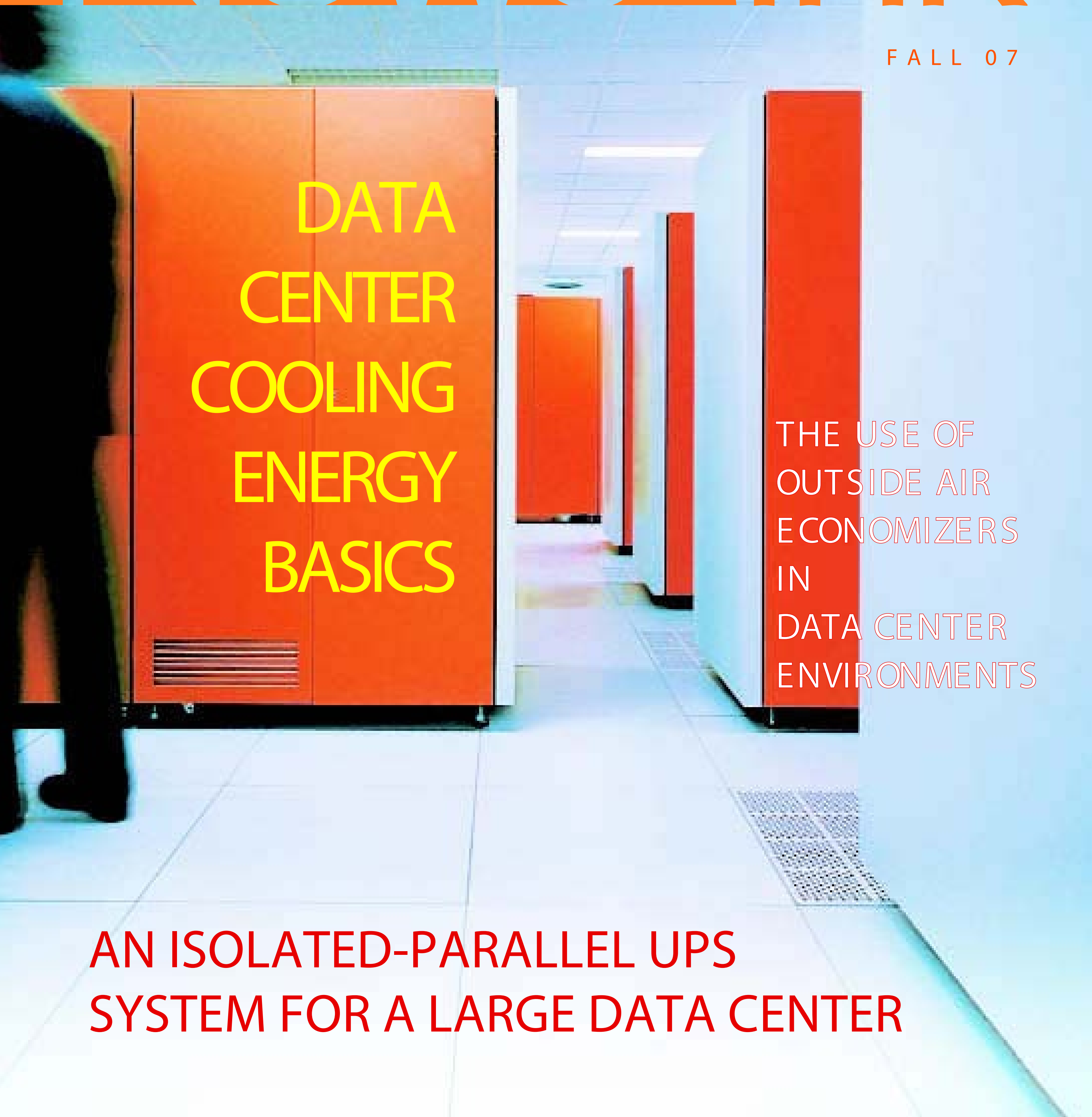
THE MAGAZINE OF 7X24 EXCHANGE INTERNATIONAL

NewsLink

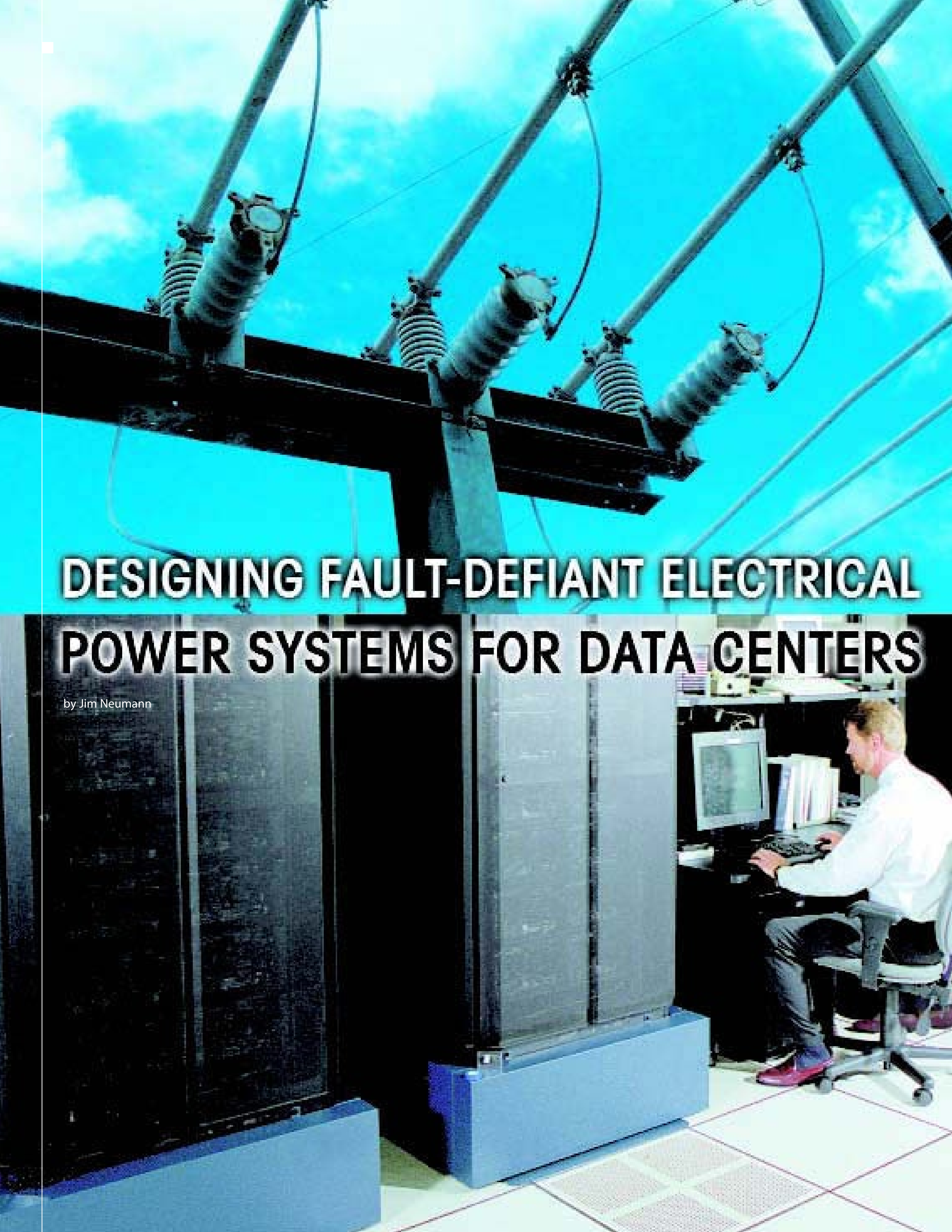
FALL 07



DATA CENTER COOLING ENERGY BASICS

THE USE OF OUTSIDE AIR ECONOMIZERS IN DATA CENTER ENVIRONMENTS

AN ISOLATED-PARALLEL UPS SYSTEM FOR A LARGE DATA CENTER



DESIGNING FAULT-DEFIANT ELECTRICAL POWER SYSTEMS FOR DATA CENTERS

by Jim Neumann

On October 4, 1957, 50 years before Fall 2007 7x24 Exchange conference, the American engineering community was jolted into the Modern Age when the former Soviet Union launched Sputnik-1, the first man-made object to orbit the Earth. Fearing that America was threatened by a shortage of engineering firepower, the U.S. Congress quickly passed the National Defense Education Act: in doing so, it also launched "The Space Race," the computer age, and millions of engineering careers.

For data center operators, a similar wake-up call came on August 14, 2003: in a mere eight minutes, the infamous Northeast Blackout moved the eastern seaboard like a tsunami, knocking out 250 power plants across a million square miles. Ten million people in the Canadian province of Ontario (about one-third of Canada's population) and 40 million people in eight U.S. states (about oneseventh of its population) were left without electrical power.

Geographically, these states are home to some of the world's most computing-intensive industries: financial services (New York), manufacturing (Ohio and Michigan), and telecommunications and healthcare (New Jersey). Of the entire FORTUNE 500, 319 companies are headquartered in states where the blackout took its heaviest toll.

For many of these companies, their data centers – even those protected by massive investments in on-site power generation, UPS, and fault-tolerant systems – were hit by surprise... and hit hard.

Figure 1: Hundreds of data center operators had a false sense of security about their resilience prior to the 2003 Northeast Blackout. (Photo: NOAA, the National Oceanic and Atmospheric Administration)

Data Centers: Corporations' Multimillion-Dollar Nest Egg

By the time power was restored, almost one-third of the nation's enterprise data centers had been affected by The Blackout. According to the data center research firm AFCOM, a survey of more than 500 data center executives found that nearly half suffered significant financial losses:

- Ten reported that they lost more than $10 million as a result of the outage
- Two reported losses of between $5 million-$10 million
- Fifteen reported losses between $1 million and $5 million
- Twenty five reported losses between $500,000-$1 million
- Fifty report losses of $100,000-$500,000
- One hundred twenty five reported loses of between $10,000 and $100,000

But for all of its notoriety, the Northeast Blackout was far from a one-time anomaly. More than 500,000 businesses and consumers experience electrical power problems every day, with the average power outage lasting two hours. The annual cost to the U.S. economy due to these power disruptions is estimated at $104–$164 billion... not including up to another $24 billion due to power quality phenomena.

According to the San Francisco Chronicle, the average public utility customer in San Francisco lost power for more than 4 1/2 hours in 2006, quoting statistics compiled by the utility and submitted to state energy regulators.

More recently, on July 25, 2007, a serious 11-hour power outage occurred in greater San Francisco. It affected major Websites –including Craigslist, Netflix, Technorati, Typepad and Gamespot – hosted at a 227,000 square foot commercial data center that was knocked offline, as much of the city blacked out.

According to Reuters, it took about five hours after the power came back on for Craigslist to be restored. And worse, according to the website Ecommerce Times, such major ecommerce sites can lose as much as $1 million per minute.

The lessons learned are simple but painful ones: interruptions in electrical power service can have costly, even catastrophic, consequences for data center operators, especially in facilities where they had invested heavily in disaster mitigation and recovery systems, and believed they were sufficiently prepared.

The Weakest Links

Why do the most intricately-designed "failsafe" systems fail?

Often times the problem begins with the design itself. Rarely is it tested to ensure its operational resilience prior to construction; and following the completion of an installation, limited tools are employed to fully test how the "live" operation might deviate from "as-designed" performance.

This seems to be unique to the data center environment: in most other mission-critical applications – the space shuttle, aircraft, shipbuilding, oil platforms, etc. – an electrical design is fully modeled and a variety of simulations are performed using advanced modeling software to test the limits of the design prior to construction.

Further, as the installation matures, undocumented changes inevitably occur; e.g. equipment or circuitry is added to a critical bus. In a crises situation, these changes result in "surprises" that can disrupt the service or recovery process.

Another common problem is that many backup systems, once installed, are no longer maintained by the experts who designed them… so maintenance procedures may not always follow those prescribed by the designer. This can be exacerbated during service actions where vital electrical components are taken in-and-out of service by personnel who may not be intimately familiar with the overall design characteristics of the electrical infrastructure.

Failure to properly maintain backup systems – or more likely, to diagnose their preparedness for changing infrastructure and loads – can have serious ramifications: 40% of all business downtime is caused by electrical power problems… and 80% of those problems are rooted in internal causes (overloading, inadequate system updates or maintenance, failing to upgrade backup systems in response to equipment changes or human error)

For example, in typical installations, an acceptable startup time for an on-premise generator is several seconds from the time of failure. During those seconds, a series of UPSs and batteries – designed to carry the load for 15 to 60 minutes – will provide emergency power and "ride-through" until the generator starts. If the generator fails start in a given period of time, this back-up (in theory) provides for a controlled shutdown of the computer equipment.

But in order to provide the desired protection, UPS units must be properly maintained. One of the most commonly-neglected components is the UPS's battery: most batteries have a useful lifetime only a few years, and they lose their ability to hold a charge gradually over that time. So if a UPS started with one hour of runtime for the connected load, after a year, it may only provide 45 minutes, and so on.

Figure 2: The cost of downtime varies by company, but the financial trade-offs to ensure maximum uptime is rarely understood until power problems take their toll

The Skyrocketing Price of Prevention

Businesses dependent on their data center operations face skyrocketing costs to build and maintain the IT systems they need to meet demand. There are five primary factors driving the spike in the cost of operating a data center:

1. Demand, Real Estate Prices are up: According to the San Jose Mercury-News, data center vacancies in the area are "in the single digits." Prime office space in downtown San Jose costs about $2 a square foot per month… but data center rents are as high as $30 a square foot per month.

2. Construction prices are up: In 2006, the typical 50,000 square foot data center would cost approximately $20 million to construct, based upon a current rate of $400 per square foot. This cost is projected to at least triple – and some say, increase to $5,000 per square – foot by 2010, meaning that today's $20 million data center construction will cost at least $60 million three years from now.

3. Consolidated, higher-capacity infrastructure: According to Gartner Inc., 94% of IT departments are in the process of either consolidating their servers, or planning to do so. Randy Mott, CIO for Hewlett-Packard, has said publicly that he plans to reduce the number of HP data centers from 85 to six… a more than 90% reduction. Thus, organizations are putting all of their "eggs" into fewer and smaller baskets… raising the IT stakes of each newly-consolidated facility. Sales of servers are estimated to increase 40% to 50% annually over the next four years, meaning that higher-density servers are being installed at a very rapid rate. Because server performance and density are more doubling every year, more and more IT resources are being squeezed into a smaller footprint… dramatically increasing the risk to reduced-site operations.

4. Utility consumption, prices are up: In the United States alone, data centers consume some where between 30 and 40 billion kilowatts of electricity annually. A typical data center consumes 15 times more energy per square foot than typical office space, and is estimated to be 100% more energy intensive, in terms of density. The more dense the server components, the more heat they generate, thus, the more air conditioning they require.

5. Obsolete infrastructure: It's not just older data centers that are being taxed by computing clusters. According to AMD, facilities built as recently as two years ago are ill-equipped to hold today's highdensity server racks, due to inherent limitations in their power utilization and cooling output.

When data centers are constructed, they are subject to a physical limit as to how much power can be delivered to the site and stored. In 2006, the typical data center was designed to handle 40 watts of power per square foot; because of increasing server density, this is projected to handle up 500 watts of power per square foot by 2010.

Similarly, data centers' air conditioning systems are limited in how much cooling they can deliver in a given space. The problem is so pervasive that through year-end 2008, heat and cooling requirements for servers will prevent 90 percent of enterprise data centers from achieving anywheres close to their theoretical server density.

The Below-the-Floor Solution

As noted earlier, the most common cause of data center failures lies in the electrical power infrastructure supporting it. Most of these failures could be prevented in either of two ways:

1. Better up-front modeling prior to construction
2. Better post-construction diligence, diagnostics, and insight

Organizations in every industry tally the cost of downtime, and make calculated decisions about the cost-benefit trade-offs of reducing power-related disruptions by investing in power systems infrastructure.

But "throwing hardware at the problem" has not resulted in failsafe facilities; to the contrary, it simply introduced different points of failure, while creating a false sense of security that causes data center operators to stop probing for potential lapses in their infrastructure.

What is needed is a more ecosystemic approach… an approach many forward-thinking organization are calling "Power Analytics." First introduced at the **Spring 2006 7x24 Exchange Conference**, Power

Analytics combines the best practices of "design" and "operations" into a single synergistic environment.

Using the original CAD model of the facility – which includes operating parameters and specifications down to the small component within the infrastructure – Power Analytics systems are able to diagnose and accurately predict the full capabilities of any given electrical power systems design.

Figure 3: The Power Analytics methodology is being adopted to protect private and public assets in a wide range of mission-critical facilities.

Further, once the system is installed and running, the "live" version of the Power Analytics system will detect the slightest variations within the infrastructure – say, a component that is slowing drifting from its normal operating parameter – and assess the downstream potential of that components deteriorating performance.

Since the performance of individual elements within power systems typically degrade with the passage time – due to aging, weathering, or stressing materials beyond their engineering limits (overloading, heat, etc.) – Power Analytics systems can identify and resolve potential problems with an extremely high degree of accuracy… typically, long before they reach the problem stage, and in sufficient time to effect repairs without business disruption.

A Crystal Ball for Power Systems Operations

While any number of "monitoring" systems have the ability to collect and manipulate data, none of these systems inherently have the capability to accurately predict and preempt disruptive electrical power problems; all require hundreds or even thousands of hours of programming following their installation in order to give the system usability that mimics "intelligence."

Power Analytics provides for an instant understanding of the systems intended design and limitations. This built-in knowledge is of extreme value by enabling the system to deliver a wide range of ancillary benefits that would standalone in their own right.

By making use of real-time operational data, this "intelligent" system makes accurate performance predictions and provides detailed operational insights into:

- Reliability – The trustworthiness of the system to perform as designed; the probability and frequency of failures... or more importantly, the lack of failures. Reliability metrics include probability of failure on demand (POFOD); rate of failure occurrence (ROCOF); mean time to failure (MTTF); and availability or uptime (AVAIL).

- Availability – The percentage of time that data can be instantly accessed, and that a system is available to deliver failure-free performance under stated conditions. The term is mostly associated with service levels that are set up either by the internal IT organization or that may be guaranteed by a third party datacenter or storage provider.

- Commissioning – The commissioning process is the first crucial step in new construction, consolidation and in some cases expansion. The model- based approach to Power Analytics can reduce the typical commissioning time from 25-50% not only providing huge cost and time reductions but for the first time guarantee that the design and the as-built drawings are in 100% agreement.

- Capacity – The storage and transaction processing capability of computer systems, the network and/or the datacenter. Capacity planning requires insightful forecasting, e.g. what if traffic triples overnight; what if a company merger occurs, etc. As a result of such the analyses and forecasts, systems can be upgraded to allow for the projected traffic or be enhanced so that they can be ready for a quick changeover when required.

- Configurability – The ease in which IT infrastructure and related systems can be maintained, upgraded, redeployed, and retired from mission-critical use. For example, understanding the effective life of specific hardware and software technology – while taking into account a systematic technology upgrade policy – allows companies to seamlessly 1) maximize the life of their technology investments, 2) incorporate new technologies, and 3) phase out older systems as their relative performance warrants.

- Energy Management – Energy management is a growing concern for mission-critical facilities, especially data centers, where the cost of powering IT equipment equals 20% of its cost. Because of the ever increasing density of servers and switches, power requirements for the largest data centers are increasing 20% per year; some data center operators report that their monthly power costs have doubled, tripled, and even quadrupled in the past 36 months.

- What-if Simulations – Power Analytics systems create virtual environment that provides an off-line, mirror image of them. It enables users to make a "freeze frame" of their real-time, current environment in order to conduct detailed "what if" simulations reflecting the present configuration of your electrical infrastructure. Such simulations include testing of real-time configuration, maintenance, repair, and other procedures, before attempting them on live systems.

Summary & Conclusions

"Fault tolerance" is no longer an acceptable measure of success in mission-critical data centers.

The technology exists today – and is proven in world-class facilities – to preempt business-disruptive electrical power problems altogether... ensuring business continuity under the most extreme operating conditions.

That's not to say that companies should abandon their investments in UPSs, on-site power generation, etc., e.g. no driver would remove the airbags from his car, simply because it comes with new collision avoidance technology.

But it's important for data center operators to know that technologies like Power Analytics hold the promise to ensure that such emergency technologies are minimized as a means to ensure day-today operational resilience.

Jim Neumann is Vice President of EDSA Micro Corp. He can be reached at jneumann@edsa.com.