Data Centers, Power, and Pollution Prevention

Design for Business and Environmental Advantage

BY

Chris Robertson and Joseph Romm

The Center for Energy and Climate Solutions

(www.cool-companies.org)

A Division of

The Global Environment and Technology Foundation

(www.getf.org)

June 2002

PREFACE

Electricity used in data centers has made headlines for at least two years and stimulated a barrage of stories in the business and technical press, some of which have shed more heat than light. This short paper is written to be understandable by the interested non-technical reader. It illuminates the intimate, volatile and high-stakes relationship between information technology—the so-called "new economy"—and the electric power system, which is firmly rooted in the industrial age.

The paper is published by the Center for Energy and Climate Solutions (CECS), a one-stop shop for helping companies and states design high-leverage strategies for reducing greenhouse gas emissions. The Center is a division of the Global Environment & Technology Foundation (GETF), a nonprofit organization dedicated to building the infrastructure for sustainable development. GETF facilitates the demonstration of new technologies and ways of doing business and helps make these ideas accessible and replicable throughout a number of sectors. We look for innovative technologies and partnerships that can significantly contribute to this goal.

Since its inception in 1998, the Center has developed best practices and high-quality case studies on corporate greenhouse gas mitigation and energy efficiency. These were published in the 1999 book *Cool Companies: How the Best Businesses Boost Profits and Productivity by Cutting Greenhouse Gas Emissions*. Since 1999, CECS has worked with the World Wildlife Fund to create the Climate Savers program, which encourages major companies to make GHG commitments. Climate Savers companies include IBM, Johnson & Johnson, Polaroid, The Collins Companies, and Nike.

Chris Robertson (crobertson@igc.org) is Principal, Chris Robertson & Associates, an energy and environmental management consulting practice focused on climate change, distributed generation, energy efficiency and renewable energy resources. Joseph Romm (jromm@getf.org) is Executive Director of CECS, author of *Cool Companies*, former Acting Assistant Secretary of the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy, and a principal with Capital E, LLC, a leading provider of consulting services for energy efficiency and distributed generation, including the high-reliability needs of the digital economy. The authors consult with several companies referenced in this paper.

We would like to thank the following people for reviewing this paper: Jon Koomey, Bill Cratty, Whit Allen and Jacqueline Abel. We greatly appreciate the assistance of Tracey Bailer and Pete O'Connor in preparing this. We are grateful to the National Environmental Trust for their support of the writing of this paper.

EXECUTIVE SUMMARY

Computers and other electronic equipment will crash at the slightest disruption or fluctuation in their supply of electricity. The power system was not designed for these sensitive electronic loads and is inherently unable to meet the technical requirements of the information economy. For data centers, which play a central role in the information economy, crashing computers cause potentially catastrophic financial losses. The same voltage sag that causes the lights to dim briefly can cause a data center to go off-line, losing large sums of money, for many hours. Data center owners and their power providers must therefore solve several related technical and economic electric power problems. These are:

- How to assure high-availability (24x7) power supply with a very low probability of failure
- How to assure practically perfect power quality
- How to manage risk while minimizing capital and operating expenses

The traditional "grid plus backup system" (which includes large numbers of batteries, diesel generators and power conditioning equipment), dominates the market. But data center information technology has advanced beyond the capacity of this traditional system to meet its needs. And the environmental impacts of the traditional grid plus backup system are becoming more and more unacceptable, including local air pollution impacts and growing concern over greenhouse gas (GHG) emissions. Data centers that can demonstrate superior environmental performance will appeal to more customers and potentially generate extra revenue in the form of saved tons of GHG emissions.

Our assessment concludes that data centers can meet society's environmental interests, and excel at their technical and business goals, all at the same time, but not with continued reliance on the conventional electric power system for their primary power supply. Such reliance bears approximately a 63 percent probability of at least one major power failure over the twenty-year economic life of a data center and worsens the data center's environmental performance.

Two important technology strategies provide data center owners the path to effective solutions for their power problems: 1) design for energy efficiency and 2) use of on-site, combined heat and power systems (CHP) designed specifically to give data centers the high-availability, high-quality power they require, with a very low (e.g. one percent) probability of failure over the data center's twenty-year life. CHP provides the primary power supply, and replaces the grid plus backup system described above.

These technology strategies are proven, cost effective, technically rigorous, and ready for widespread adoption. Data centers that implement these technologies can expect to reduce total energy use, GHG emissions, and air pollution by 50 percent or more. Data center capital cost and annual operating cost savings of about 20 percent each can be achieved with CHP. For the electric utility industry these technologies represent a new opportunity to meet the needs of their new-economy customers.

In this paper we answer these questions:

- What are data centers?
- How much energy do they use, and what are the trends for future energy consumption?
- What are the environmental impacts of data center energy use?
- What is high-availability, high-quality power, how is it measured, and why is it needed?
- How can a data center's power supply be designed to be available 99.9999% of the time?
- What are the data center's risks if their power supply fails?
- What is the role of energy efficiency and on-site CHP primary power supply?

WHAT ARE DATA CENTERS?

Data centers house the computers, servers, switches, routers, data storage devices and related equipment used to operate the digital economy. They can be relatively small facilities owned by a single company and dedicated to processing its data alone. At the other end of the scale, they can be huge facilities where many companies, such as Yahoo or eBay, can rent floor space containing wire cages filled with racks to house their servers and other equipment. Or the facility may have a single owner who provides managed information technology (IT) services and acts as an application service provider (ASP) on an out-sourced basis. These facilities can go by many names while pursuing varied business models: internet hotel, data farm, data warehouse, co-location facility, server farm, telco hotel, internet service provider, ASP, corporate data center and others.

Whatever their name or business model, these facilities all share certain common attributes. Inside the data center are racks of electronic equipment wired together and connected to fiber-optic cable networks. These equipment racks are closely spaced, and equipment densely packed in order to maximize the amount of electronic equipment that can be installed.

To operate this equipment, data centers require large quantities of premium electric power that meets the most stringent quality and availability levels. Large capital investments ($400 - $800 per square foot) are required to provide power, air conditioning and ventilation. How much power, and how it can be reliably and continuously supplied, has been a matter of intense interest among utilities, data center owners, their engineering firms, suppliers, investors, analysts, environmentalists and other stakeholders.

Data center discussions are often couched in terms of something per square foot (i.e. profit, capital expense, electric power, etc.). This can be confusing depending on what parts of the data center one is discussing. In her excellent Masters Degree thesis, Jennifer Mitchell-Jackson from the University of California at Berkeley describes how space is allocated in a data center.

"In an average hosting facility, approximately 50% to 60% of the facility's footprint is designated for computer equipment. … The main computer room area is commonly referred to as the net floor space. This area is also often called 'raised floor area' due to the fact that the floor is usually elevated so that cool air can be blown in from underfloor air ducts. Within this area, 20% to 33% of the floor space is usually covered with rows of equipment racks or cabinets (i.e., enclosed racks). The low coverage of this area is due, in part, to building codes that require three foot aisles between racks. Approximately 50% of the area in the computer room is inactive aisle space or service clearance area. The remaining 20% is usually for support equipment such as computer room air conditioning units (CRAC units) that must currently be located near the computers."[1]

POWER DENSITY

Design power density is a measure of the electric supply capacity installed to power the data center. Actual power density is the instantaneous electric demand required by the facility when it is operating. The common term to discuss power density is Watts per square foot (W/sf). As Mitchell-Jackson wrote, this can be a confusing metric depending on whether one is referring to "1) a single piece of computer equipment, 2) a single rack, or 3) the footprint of all of the racks excluding aisle space, 4) the footprint of the central computer room including aisles between racks but excluding exterior mechanical rooms and office space, 5) a single data center floor within a multipurpose building or 6) the footprint of the entire building." She went on to define the total "computer room power density" that includes all computers and the associated equipment required to power and cool the computer room.

Data centers have been requesting electric service capacity from local utilities at between 75 and 300 W/sf of computer room power. Those who have measured the actual energy demands put the actual present power density for all equipment at about 40 – 50 W/sf.[2] This is on the order of 10 times the amount used by a typical commercial building.

Why is there such a difference between design and actual power density? There are many reasons, including engineering conservatism and over-design. Based on our conversations with data center facility staffs, we understand that owners tend to be risk averse. They don't want to be caught short of capacity if future equipment power requirements rise. They see other providers guaranteeing high power and they can't afford to offer less.

Several competing trends make future data center power density difficult to forecast. Consider these examples. Some will reduce power density and some increase it.

- **Low energy servers.** Vendors like RLX Technologies are using low power chips and innovative designs to reduce the energy demand of servers by more than 80 percent. IBM has announced new server technology with similarly large reductions in energy use.[3]

- **Miniaturization.** RLX can pack more than 5 times as many servers in the same space as their competitors. IBM's offering replaces racks of servers with one small mainframe.

- **Increasing equipment density.** "Ultra-dense" severs are now coming to market. Racks that once held 42 servers can now hold 336. Even with lower power servers this can drive data center power density higher.[4] The Uptime Institute expects "server racks to trend from current rack power densities of 600 W/ft^2 to 1600-1800 W/ft^2 by 2010. Telecommunication frames, which are typically more energy intensive than server racks, may reach product densities of 7000-9000 W/ft^2 by 2010."[5] This would create a very demanding requirement for heat removal.

- **Increasing Internet traffic.** Internet traffic continues to grow, with one study from mid-2001 saying it is doubling every six months.[6]

- **Increasing data intensity in Internet messages.** Streaming video is much more data intensive than simple text messages. Compare a simple text message with data requirements of 20 or 30 kilobytes to a 15 second streaming video with 1220 kilobytes required.[7]

- **Business model variations.** Corporate data centers are more likely to adopt efficient technologies like IBM's because they control both the facility and the electronic equipment purchase decisions. Co-location facilities, which lease rack space to 3rd party service providers, have practically no control over what electronic equipment is installed, and must design for load flexibility.

- **Design for Energy Efficiency.** The physical facilities that support data centers, like the heating ventilating and air conditioning (HVAC) and electric supply system components, can be designed to be substantially more energy efficient. This is discussed in greater detail later.

DATA CENTER DEMAND FOR ELECTRIC UTILITY INFRASTRUCTURE

From 1999 to early 2001 electric utilities in the top 25 to 30 US markets observed huge increases in electric service requests from data centers. A few examples illustrate this point.[8] Austin Energy reported 106 megawatts (MW) of new requests; Puget Sound Energy had initial requests for over 1000 MW; New York more than 500 MW; Sacramento 65 MW; Pacific Gas and Electric over 1300 MW. In reality, very little of this requested capacity was actually installed before the dot-com crash, giving both utilities and the data center industry time to reassess their strategy. For example, just 6 MW of data center load has materialized for Austin Energy.

Utilities had both financial and engineering constraints that slowed the expansion. From an engineering perspective, it was all but impossible to meet many of the requests in the time frame sought by some data center developers. Often utilities were asked to supply large distribution capacity increases in three months, where it would typically take two years to order and install the substation transformers and other equipment. Some utilities flatly discouraged data centers from locating in areas of their system that could not be reengineered to handle the huge expected loads.

Financially, many utilities (and other analysts) wondered whether these new mega-customers would be around long enough to pay the infrastructure upgrade costs. So they imposed hefty upfront infrastructure charges. Some data center developers had assumed in their business plans that the utility would just be there, and were completely unprepared to pay these fees on top of their other infrastructure costs.

The lack of flexibility in the capacity expansion plans of these facilities contributed to the problem. If a utility and data center could start small and add increments, or modules, of power supply as needed, then flexibility would be improved. But such modularity is not easily achievable when one asks a utility to build a new substation to serve a large new electric load for a 20-year project. The utility and the customer will want to install the required capacity at the front end, because of the long lead times to get additional transformers and distribution system equipment.

ENVIRONMENTAL IMPACTS

The projected build up of new data center electric loads led directly to forecast increases in energy-related air pollution, both globally and locally.

Greenhouse gas (GHG) emissions such as carbon dioxide from fossil fuel-fired electric generators are a major cause of global climate change. The forecast increases in data center electric demand implied a growing source of GHG emissions. This is cause for concern. The international community is attempting to reduce GHG emissions amid broad scientific and international consensus that climate change is a major threat and must be taken seriously.

Air pollution from diesel generators causes environmental health problems. Each planned data center includes as many as two or more times redundant backup diesel generator capacity, to use when the electric grid is unstable or unavailable. Diesel generators used for emergency backup power supply are essentially unregulated.[9] They are a notorious source of very high levels of damaging air pollutants, including soot, nitrogen oxides (NOx), which form ground level ozone (smog) during hot sunny weather, and others. Coincidentally, the electric grid is most unstable and at risk of failing during hot periods of peak air conditioning demand.

San Jose, California alone has more than 1000 MW of emergency backup diesel generators. During the summer of 2001 there was concern that a big brown cloud of diesel air pollution would be present over many cities during the hottest days. The environmental regulatory community responded with alarm bells to this forecast data center growth with its large dependence on emergency diesel generators. It is likely that increased regulation of backup diesels will result.

Because of concerns over pollution and strains to electric utility infrastructure, many companies, organizations and academic analysts have begun to examine how to improve energy performance in data centers, so as to reduce energy use and consequent air pollution emissions. Among the strategies that have emerged are design for energy efficiency and on-site combined heat and power (CHP). These strategies address the data centers' environmental impact issues. Perhaps more importantly to data center companies and electric utilities, they make solid contributions to technical and financial risk management and bottom line profitability, greatly improving their chances for widespread adoption.

POWER QUALITY AND AVAILABILITY

Electronic equipment operates with continuous flows of electrons. Tiny disruptions in the current or voltage of the electric power supply will corrupt data or crash computers. For data centers, these fluctuations can cause large financial losses, so they require high-availability and power quality 24 hours per day, 7 days per week, for twenty years.

To address electronic equipment's sensitivity to low-quality power, the Computer Business Equipment Manufacturer's Association (CBEMA) developed the so-called CBEMA Curve more than two decades ago. This is a power quality specification for electric power supply used to power computers. It has been adopted as Standard 446-1987 by IEEE, the Institute of Electrical and Electronics Engineers.

Power quality from the local electric utility is not adequate for electronic loads. Voltage fluctuations, surges, spikes, sags, and momentary outages occur frequently. Examples are commonplace. A neighbor's large motor starting up can cause a sag in the power system voltage large enough and long enough to trip off a computer or server. Lightning can cause a spike in the voltage, with the same bad effect on electronic equipment. Longer outages occur, but less frequently. As a result, from a computer's perspective, the electric utility system is only "available" less than 99% of the time.

To address the power quality and availability issues, data centers traditionally rely on the electric grid plus backup power supply. They typically install 100 percent (also known as 2N) redundant power systems consisting of dual utility feeders from separate substations connected to on-site power-conditioning equipment. This includes uninterruptible power supplies (UPS), large strings of storage batteries and emergency diesel generators. In this scheme, the utility power is input to the UPS, where it is conditioned to meet the CBEMA specification and then supplied to the electronic equipment. When momentary outages occur, the batteries take over and supply power until the utility power returns. If the outage is not brief, the diesel generators start and are switched over to supply power to the facility.

It's not just the electronics that have to operate continuously. The air conditioning and ventilation system must also run, because the heat generated by the electronics is so intense that the equipment will over-heat and malfunction if it is not continuously cooled.

MISLEADING METRIC: "HIGH 9s" AVAILABILITY

"Availability" is a statistical term. It means the "fraction of time that a repairable system is available for use."[10] Data center owners desire their power supply to be available 99.9999% of the time, or "six 9s." A system available 99.99% of the time is a four 9s system, and so forth.

The appropriate engineering method for calculating the availability of repairable complex systems such as a data center is Probabilistic Risk Assessment, or PRA. The science of PRA has been highly developed over the past 50 years and is used extensively in aircraft, defense, insurance and the nuclear power industry. *PRA is used to a limited degree in the data center industry, and in our opinion should become the standard technique for evaluating claims of high-availability power and HVAC systems in data centers and other mission-critical facilities.*

Recent work by MTechnology, a leading firm in the field of risk assessment, used PRA to study typical data center power supply scenarios for one of their clients. They concluded that *at best* the traditional system designs, (that is, a 2N redundant system) can *theoretically* achieve five 9s availability (99.999%), but in practice only achieve four 9s (99.99%) or less.[11]

That is, the conventional electric power grid plus UPS, batteries and backup diesels is only a four 9s system. The reason for this is that single points of failure, switching failure (though rare), and potential cascade failures put an upper limit on availability of the traditional system. Adding more equipment, components and levels of redundancy actually increases the probability of failure because more switching is required, more single points of failure are introduced, and the potential for cascade failure increases.

A single point of failure in a system is something that, if it fails, brings the whole system down. Cascade failure occurs when a failure event happens in interdependent subsystems. For example, when one of several operating emergency backup generators throws a rod, the rest of the generators cannot meet the load, and the whole system fails. Failure to start, and to keep running once started, are relatively common events for emergency backup diesel generators. A recent study of diesel generator performance in the nuclear industry over many years concludes that diesel generators fail to start about one percent of the time and almost 5 percent fail within 8 hours running time.[12]

These failure modes pose fundamental constraints to high nines availability in the traditional grid plus backup power supply system. When the electric grid power fails, the multiple redundant backup generators must start up and then synchronize their rotational frequency, establish parallel operation with each other, switch their output to the data center's power distribution system, and run continuously until they are no longer required.

There is a huge difference in probability of failure over a twenty-year project life between a four 9s and six 9s system. As the figure below shows, six 9s systems have a 1 percent probability of at least one major failure in the 20 year economic life of a data center. This is what data center owners would prefer to have, because their data storage systems are engineered to a six 9s standard. If their power supply doesn't match this, then the power system becomes the weak point in the system and compromises the electronic equipment's high availability. Four 9s systems have a 63% percent probability of at least one failure over data center's 20 year life. This is because cascade failure, switching failure and single points of failure cannot be engineered out of the traditional power and HVAC system.

A recent example demonstrates that failure of the grid plus backup generators can have serious negative consequences for data centers and their customers. Pacific Gas and Electric had a fire in an underground transformer vault that served a data center. The data center's backup power supply system failed. The result was that the data center went down for hours. Yahoo and other e-commerce sites went down with it, and these companies incurred substantial financial losses in a matter of a few hours.

To our knowledge, only one data center, owned by the First National Bank of Omaha, has been designed and built using PRA to specify the power system design, maintenance protocol and spare parts inventory requirements, as well as to evaluate the installation post-construction. Its primary power supply is not from the grid; rather it is from a novel, on-site combined heat and power system.

FNBO is a large independent credit card processing facility. Its customers had experienced substantial losses from power outages in FNBO's previous facility. When FNBO built its new Technology Center in 1999-2000 it specified six 9s availability in the power supply contracts and required 3rd party PRA to document performance. FNBO uses this six 9s availability as a lead feature in its marketing campaign, and has grown its market share as a result.

It is common to hear claims of six 9s (or even greater!) performance by data centers and vendors to the industry, because the cost of failure is so great and the stakes so high. These claims should be examined with great caution and skepticism. Any investor or customer of a mission-critical facility with large financial exposure and dependence on high nines performance should ask to see the mathematical PRA analysis that provides the basis for such claims. The analysis should be transparent, performed by a third party, and replicable by competent engineers.

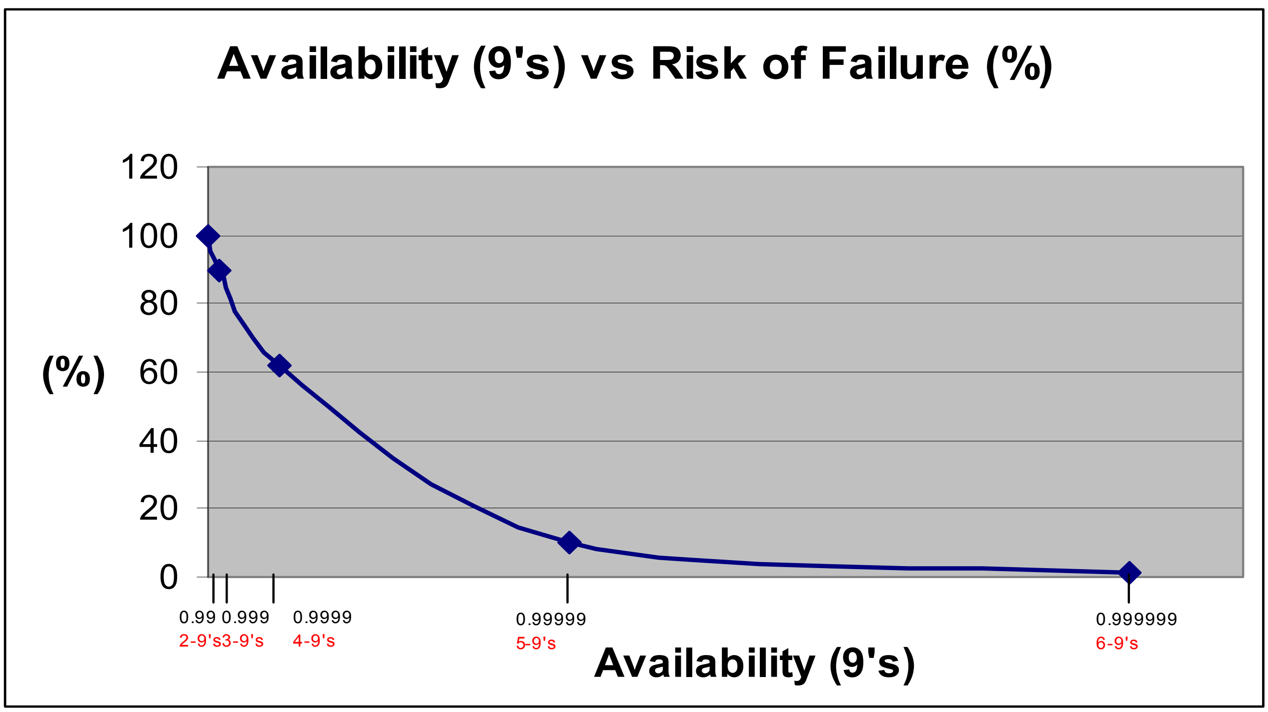

Figure courtesy of MTechnology

THE REAL DOWNTIME RISKS

Claims of high nines performance are often associated with statements like “Our six nines system will only fail for 32 seconds per year.” This is highly misleading. A study by PlanetTech and MTechnology for the Massachusetts Technology Consortium explains:

> “The standard, but inaccurate means of explaining high availability is to take the corresponding unavailability and multiply it by the amount of time in one year. As there are approximately 8760 hours in a year, a table such as the one shown below results:”

Nines	Availability	Unavailability	Time unavailable per year	
"one nine"	90%	10%	876	Hours
"two nines"	99%	1%	87.6	Hours
"three nines"	99.9%	1.0.E-3	8.76	Hours
"four nines"	99.99%	1.0.E-4	53	Minutes
"five nines"	99.999%	1.0.E-5	5.3	Minutes
"six nines"	99.9999%	1.0.E-6	32	Seconds

Table: Typical and "inappropriate" means of determining Unavailability

> “This table is inappropriate for virtually all Internet related industries, and inaccurate for any values higher than “three nines.” How can such a simple calculation be inaccurate? The key is in the definition of availability. Remember that the term “six nines” refers to electricity that is available and suitable to power computer loads so as to keep them operational 99.9999% of the time. The problem that leads to the erroneous table is that electrical power that deviates from specified norms for more than a mere fraction of a second will cause the thousands of servers, routers, switches, disk arrays, and other equipment in an Internet industry to crash, hang, reset, or otherwise fail. When they fail, it takes a large effort by the system operators to restore normal service. The task is not simply rebooting computers, but restarting processes, re-establishing high-speed communication links, repairing and synchronizing large databases that may have been corrupted, and so forth. Typical Internet centers hit by even a 0.020 second “blink” in the electrical power may require 16 hours or more to restore normal operations.
>
> “The fallacy in the table above is that there is no such thing as a 32 second, or 5-minute outage in an Internet industry. Systems designed according to the assumptions consistent with the table will fail, rather spectacularly, to achieve sufficient availability. The dot-com meltdown is littered with the corpses of companies that did not understand the consequences of momentary power outages on their operations. The classic example is eBay, which lost not only revenues but also a large fraction of their regular vendors after a 72-hour outage. Not only did the loss of vendors to competing auction sites cause long-term financial harm to eBay that far exceeded the loss of 72 hours of operating revenues, but the stock market responded with a substantial, long-term devaluation in eBay’s stock price.”[13]

ENERGY EFFICIENCY AND COMBINED HEAT AND POWER

Improving data center's energy efficiency and reducing their environmental impacts is dominated by three major energy related choices: electronic equipment, air conditioning systems and primary power supply.

Energy-efficient servers: The server industry is rapidly evolving with respect to the efficiency of their products. RLX Technologies initiated the energy efficient server market using chips from Transmeta Corporation in the spring of 2001. In less than six months major industry players were working to improve server energy efficiency. In late 2001, both IBM and Intel announced initiatives to improve the energy performance of data center electronic equipment.

In October 2001, IBM announced "its launch of a new low-power computing research effort, aimed at achieving energy efficiency in information technology. Coordinated out of its research lab in Austin, Texas, the research effort will accelerate the development of energy-efficient chips, computers, servers, and other components. A complementary consulting service will help customers lower their information technology energy use." EPA has recognized IBM for its energy-efficient eServer z900 mainframe. According to the EPA, "IBM not only applied its semiconductor technologies such as copper wiring and silicon-on-insulator (SOI), but reinvented the mainframe to handle the unpredictable demands of e-business, allowing thousands of 'virtual servers' to operate within one box. Companies that in the past required hundreds of individual UNIX-based servers now can potentially save significant energy and dollars by using one eServer z900 to accomplish the same task."[14]

Six weeks later Intel announced, "a new series of products for the emerging category of low-power, space-saving 'ultra dense' servers. 'The need for ultra dense servers is rapidly emerging, particularly in Internet data centers where there's a high priority on lowering power usage and real estate costs,' said Richard Dracott, director, Intel Enterprise Products Group."

According to Intel, their new products deliver "server-specific reliability and performance features [and] offer the most compelling performance [in power density] - an increasingly important factor for IT purchasing decisions and key performance metric for enterprise data centers and Internet Service Providers."[15]

Energy-efficient air conditioning systems: Data centers use a lot of cooling capacity because all the electricity used to power the electronic and other equipment is converted to heat and must be removed. A 60,000 square foot data center will require between one and four thousand tons of cooling capacity, depending on its power density and design. The data center air conditioners run all year, day and night.

Data center air conditioning systems are typically air-cooled direct-expansion roof top units (RTU). They produce cold air that is distributed by computer room air conditioners (CRAC units) throughout the data center to ventilate and cool the servers and other equipment. This technology is being improved.

In September 2001, RTKL came out with a new direct air-cooling system for data centers. In the same month, Sanmina Enclosure Systems announced a new water-based cooling technology for data centers, though it remains to be seen whether data center owners will want water flowing around the cabinets holding their servers.[16]

Data center developers seek to minimize capital expense, and RTUs are the air conditioner technology of choice. RTUs have low capital costs but they are much less efficient as compared with the alternative, which is a central HVAC plant incorporating water chillers, pumps, and cooling towers. The metric used to measure the efficiency of air conditioners is kilowatts per ton of cooling capacity, or kW/ton. RTU efficiency is on the order of 1.3 to 1.4 kW/ton. Typical central HVAC plants are designed to operate at about 0.85-0.9 kW/ton.

One vendor to the data center industry, Sure Power Corporation, has developed a central chilled water plant for data centers that delivers six 9s availability and operates at 0.29 kW/ton. This performance ranks it as probably the most energy-efficient HVAC system that could be used in a data center.[17] The plant is modular; this means cooling capacity can be added or subtracted without affecting the system operation or its high availability. They use independent PRA to calculate both as-designed and as-built availability claims, and to plan and implement maintenance practices. It is designed to operate as part of an on-site CHP system.

Energy-efficient, on-site, combined heat and power: On-site electric generators are used in a CHP system as the primary power supply. Waste heat from the generators is used to provide hot water, steam or chilled water; hence the name, combined heat and power. Data centers require relatively constant quantities of electricity and air conditioning, which matches up well with CHP capabilities.

CHP systems are usually more than twice as efficient as the conventional electric power grid, for three reasons. The generators can be more efficient than the average of the electric grid. Waste heat can be used on site. And the normal losses from electric transmission and distribution are avoided. Because of these energy efficiency aspects, CHP is an important technology for reducing greenhouse gas emissions caused by electric power generation.

In April 2002, Verizon was awarded a DOE grant through a program aimed at supporting distributed energy resources in applications for data processing and telecommunications. Verizon was awarded up to $3 million as part of its Central Office of the Future Project. Verizon will use multiple fuel cells and reciprocating engine generators to power a large central office facility in New York, where waste heat will be utilized for CHP. The project will increase understanding of controls for multiple DER units; as well as, utilize low-grade heat for CHP benefits.

Sure Power Corporation was also awarded a DOE grant in this program, receiving up to $2.2 million to support the development of an innovative DER/CHP system having six 9s availability. A system design that optimizes reliability, cost and power quality will be tested as a retrofit to a Cable & Wireless/Exodus Internet data center on the West Coast of the U.S. Sure Power Corporation has developed CHP technology for fully integrated, six 9s power and HVAC systems. Their designs are guided and performance documented by independent PRA. Design requirements include the following:

- Use on-site primary power generation in a system configuration that meets high availability requirements of the data center industry
- Generators are energy efficient with very low air pollution emissions to minimize both local and global environmental impacts
- Use waste heat from power generation to produce air conditioning in a high efficiency and high-availability central HVAC system
- Eliminate batteries, UPS and backup diesel generators
- Reduce capital and operating costs (by 20 percent or more) as compared to the conventional electric grid plus UPS, batteries and backup diesel generators
- Use modular and scalable design so that components can be added or subtracted at any time to meet the risk management requirements of the electric utility and data center industries.[18]

MODULAR AND SCALABLE SYSTEMS IMPROVE RISK MANAGEMENT

In the face of uncertainty about data center energy intensity, using the traditional system of the electric grid plus UPS and backup diesel generators represents a source of financial risk to both the IT and electric utility industries. For example, if a data center designer specifies 200 Watts per square foot on the expectation that the future power density will reach that level, then power and HVAC capital investments must be built to that specification, with little room for modular expansion; substations must be built, transformers purchased and installed, power lines run, and significant parts of the facility's electrical and HVAC systems installed to meet the expected load.

But if the electric load growth is slow, or new technology reduces power density to a fraction of the installed capacity, then substantial capital investment has been sunk. This increases the cost for the data center, making it less competitive and perhaps putting its finances in jeopardy.

As we've seen, though, there is great uncertainty about future power density and some observers expect it to increase. Suppose a data center installs sufficient capacity for, say, 70 Watts per square foot, and future energy requirements exceed that expectation. The utility has installed capacity to meet the 70 W/sf design, but now must be asked to come back and install additional capacity. This could be a time-consuming, costly proposition, probably more than a year if new transformers or other substation upgrades are required. To make matters worse, the data center must then allocate additional space to batteries, UPS and backup generators, which reduces space available to rent. Costs go up and revenue declines, which is not a happy scenario.

On the other hand, if one selects an integrated, on-site power and HVAC system which is designed at the start to be modular and can be scaled up or down as need be, then the capital investment in electric and HVAC system capacity can be matched to the electric and HVAC loads as they evolve over time. If the load grows rapidly then more capacity can be added quickly. If the loads grow slowly, then scarce capital has not been tied up in useless infrastructure investments. This can have the salutary effect of matching capital investment more closely to cash flow, reducing debt requirements and improving profitability.

For electric utilities, widespread adoption by data centers of distributed combined heat and power technologies would have significant benefits. CHP decentralizes the power generation system, moves the generation closer to where it is needed and reduces the need for expensive transmission and distribution upgrades to serve the new data center loads. It improves energy efficiency and environmental performance. It permits discrete, timely and correctly sized increments of generation to be added or subtracted where and when needed. It reduces the risk that future region-wide power disruptions will be blamed on data centers and the Internet. And it provides electric utilities with a new customer service option, six 9s power and HVAC services, which they are presently unable to deliver.

STRATEGY SUMMARY: HOW TO RECONCILE DIVERSE INTERESTS

Data centers can solve their technical, economic, risk management and environmental electric power problems, but not with continued reliance on the conventional electric grid, even with the addition of redundant batteries, UPS and backup diesel generators. In practice this achieves a 63 percent probability of at least one failure over the twenty-year economic life of the data center. The high-availability and power quality that data centers require can be provided by on-site, primary, combined heat and power systems. CHP can achieve a one-percent probability of failure over the data center's 20-year life. Third-party verified probabilistic risk assessment is required to assure that high-availability is truly achieved.

Economics favor the CHP system. As compared to a traditional 100 percent redundant system of the grid plus UPS, batteries and diesels, a CHP system has lower capital cost and lower annual operating cost, both on the order of 20 percent.

How much energy savings and environmental improvement can be expected from best practices? This is a difficult answer to pin down, because of uncertainty about what the power requirements will be for a "typical" data center. However, we can say some things with relative certainty.

Server and other electronic equipment power density is a wildcard. Trends in lower power chips are real, as are trends to densely pack servers into racks. How these trends interact will be fascinating to observe, but are beyond the range of our ability to forecast.

Combined heat and power systems are substantially more energy efficient than any other way of providing high-availability power and HVAC services to a data center. Given a data center of any assumed power density, CHP can produce energy savings of about 50 percent.[19]

Enhanced ventilation technologies such as direct cooling of the server racks appear to offer a significant efficiency as well as technical improvement. These technologies can further reduce data center energy use.

Environmental improvements are even greater than these energy savings imply. Local air pollution from power generation is a source of serious health problems. CHP technology eliminates emergency diesel generators and their associated air emissions from the data center. The diesels are replaced with clean burning, natural gas-fueled generators, which operate continuously. Outfitted with the best available control technology, they reduce air pollution emissions by 99 percent and meet the most stringent local air pollution regulations in the US.

Greenhouse gas emissions are similarly reduced. In a study by the authors, the GHG emissions of a traditional data center design were compared to the same data center using onsite combined heat and power with a highly efficient HVAC system design. The result was a 60 percent reduction in energy related carbon dioxide emissions.

One last issue deserves mention. Data centers became the poster child for those who believe the Internet is an electricity hog because their designers have been asking utilities for so much power, and because, before the recent telecom collapse, their were literally dozens proposed for every major city in the country. As it turns out, this view of the Internet, advanced by Mark Mills and Peter Huber, was based on flawed analysis, as reports by Lawrence Berkeley National Laboratory and A. D. Little have shown. The Internet is not a big electricity consumer nor is it likely to be in the future.[20] As for data centers, using 50 watts per square foot for their power density and a relatively optimistic estimate of the existing square footage of data centers in the country, they currently consume just 0.12% of U.S. electricity. Even using pre-telecom-collapse estimates of data center growth, they would add no more than 0.1% to annual electricity demand growth. In all likelihood, they will add far less than that to U.S. electricity demand growth.[21] If they are designed with energy efficiency in mind, as we have described above, their demand on the electric power system and environmental impacts will be even further reduced.

CONCLUSION

In this assessment we have examined the relationship between the information technology and electric utility industries. They represent two ages – the new digital computer-based economy and the old industrial economy. We have shown that their separate interests create several related problems that cannot be adequately solved by the conventional technology approach.

The first two, how to assure high availability and high quality power, are absolutely required by data centers. The third, how to design power and air conditioning systems to manage risk while minimizing capital and operating costs, is highly desirable to both utilities and data centers. The fourth, how to minimize environmental impacts of power generation, is a pressing social issue that can be made either better or worse by technology choices.

Fortunately, a few key technologies are finding their way into the data center industry. These technologies are design for energy efficiency and combined heat and power. They provide a foundation on which utilities and their high technology customers can build effective technical systems, manage risk, and meet the environmental protection needs of society.

ENDNOTES

[1]Jennifer D. Mitchell-Jackson, "Energy Needs in an Internet Economy: A Closer Look at Data Centers," M.S. Thesis, Energy and Resources Group, University of California, Berkeley, 2001, pp 16-17. This can be downloaded as a PDF file at http://enduse.lbl.gov/projects/infotech.html.
[2]Ibid. Commonwealth Edison found a similar result. See Thomas Callsen, *Transmission & Distribution World*, "Internet Hotels—A Big Draw" (February 2001). Personal Communications with Thomas Callsen. See also, Beck, Fred, "Energy Smart Data Centers: Applying Energy Efficient Design And Technology To The Digital Information Sector," Renewable Energy Policy Project (REPP), November 2001 [Hereafter REPP].
[3]RLX Technologies, "Product FAC," The Woodland, TX, 2002. www.rlxtechnologies.com/products/products_product_faq.php.
IBM Launches 'High-Powered' Initiative for 'Low-Power' Products and Services, Press Release, October 1, 2001. www-916.ibm.com/press/prnews.nsf/jan/5328A987B028E83085256AD8004A3972
[4]Rich Miller, "Smaller Servers, Larger Loads," *CarrierHotels News* online, July 31, 2001. www.carrierhotels.net/news/July2001/imnpower0731.shtml
[5]REPP, p. 13.
[6]Rich Miller, "Is Internet Usage Surging?" *CarrierHotels News* online, August 16, 2001. www.carrierhotels.net/news/August2001/netgrowth0817.shtml
[7]Data from Morgan Keegan & Company, cited by Nordham et al., "Delivering Energy Services to Internet Hotels and Other High-Density Electronic Loads: Part I: Structure of the HiDEL Industry," E-Source, a unit of Platts, 3333 Walnut Street, Boulder, Colorado 80301-2515, E-mail esource@esource.com.
[8]REPP, p 7.
[9]Emergency generators are permitted to operate without emissions controls for up to 50 hours per year in California, and up to 876 hours per year in Texas, both states with severe NOx pollution levels.
[10]Stephen Fairfax, MTechnology, Inc., "High Availability: Methods, Myths and Mistakes," Keynote Presentation, 7x24 Exchange Fall Conference, "End to End Reliability: The Future," November 2001.
[11]William Cratty, Chief Technology Officer, Sure Power Corporation, personal communication, February 7, 2002.
[12]Grant, G.M., et. al., "Emergency Diesel Engine Generator Power System Reliability 1987-1993," Idaho National Engineering Laboratory, INEL-95/0035, February 1996.
[13]"A Technical and Market Assessment of Premium Power in the Haverhill Cyber-District," Prepared by Planet-TECH Associates (with MTechnology), January, 2002
[14]IBM Press Release, op. cit.
[15]"Intel Releases New Line Of Low-Power Server Products," Intel press release, Nov. 13, 2001. www.intel.com/pressroom/archive/releases/20011113comp.htm
[16]REPP, pp 13-14
[17]William Cratty and Whit Allen, "Very High Availability (99.9999%) Combined Heat and Power for Mission Critical Applications," Sure Power, Danbury, CT, November 2001. www.surepowersystem.com
[18]Cratty and Allen, op. cit., p 9-10.
[19]William Cratty, Chief Technology Officer, Sure Power Corporation, personal communication, February 7, 2002.
[20]Peter Huber and Mark Mills, "Dig more coal—the PCs are coming," *Forbes*, May 31, 1999, pp. 70-72. The underlying analysis is Mark Mills, *The Internet Begins with Coal: A Preliminary Exploration of the Impact of the Internet on Electricity Consumption*, The Greening Earth Society, Arlington, VA, May 1999. Various critiques by Jon Koomey and others of Lawrence Berkeley National Laboratory can be found at http://enduse.lbl.gov/Projects/infotech.html. See also Kurt W. Roth et al, *Energy Consumption by Office and Telecommunications Equipment in Commercial Buildings*, A.D. Little, Cambridge, MA, January 2002.
[21]Mitchell-Jackson, op. cit., and personal communication with Jon Koomey.