An ACEEE White Paper: Overview of Data Centers and Their Implications for Energy Demand

Prepared by
Elizabeth Brown, R. Neal Elliott, and Anna Shipley

September 2001

1001 Connecticut Avenue, NW, Suite 801, Washington, DC 20036
(202) 429-8873 phone, (202) 429-2248 fax, http://aceee.org

Acknowledgements

The authors express appreciation to the Office of Power Technologies at U.S. Department of Energy for support that enabled us to prepare this report. In addition, we would like to acknowledge the assistance of numerous individuals that contributed to the collection of the information contained. In particular we would like to acknowledge: Robert Bryce of *Interactive Week*, Roger Duncan of Austin Energy, Chris Robertson of Chris Robertson Associates, Jonathan Koomey and Jennifer Mitchell-Jackson of Lawrence Berkeley National Laboratory, Grant Duhone of PG&E, and David Van Holde formally of Seattle City Light and now of FT Energy.

ABSTRACT

The field of high tech building studies is relatively young, but given the rapid increase in these buildings with the booming economy of the late 1990's there is much opportunity for discussion and study of growth focusing on the industry's energy use and efficiency opportunities. This paper reviews what we currently know about the market for data centers, including recent market changes and future projections, current power need, and the way in which that need is met. After the review, we list possibilities for energy efficiency, stressing the use of combined heat and power as an alternative to diesel generators for reliable power.

DATA CENTERS IN CONTEXT

The field of high tech building studies is relatively young, but given the rapid increase in these buildings with the booming economy of the late 1990's there is much opportunity for study of growth focusing on the industry's energy use and efficiency. The opportunity was highlighted by concerns raised by utilities, which began receiving numerous requests for new, large-capacity loads with very short lead times. ACEEE became aware of the issue in early 2000, when they were approached by several utility contacts concerning the topic. This interest lead ACEEE staff to organize an informal session on the topic at the *ACEEE Summer Study on Energy Efficiency in Buildings* held in Pacific Grove, CA in August of that year. This gathering provided one of first opportunities for various utilities across the country to share their experiences with what was a fast-growing market.

The current state of the economy has dramatically slowed the rate of growth of data centers. However, with the amount of information exchange continuing to increase, and increasing security and data redundancy issues, data centers will remain a staple of communication and data storage.

THE DATA CENTER MARKETPLACE

In order to understand energy efficiency opportunities in data centers, an understanding of the current data center industry is necessary. Emerging in the late 1990's, data centers are locations of concentrated Internet traffic requiring a high-degree of power reliability and a large amount of power relative to their square footage. Typically, power needs range from 10-40MW per building, and buildings are typically built in clusters around nodes in the Internet fiber-optic backbone. During the development boom in 1999 and 2000, projects averaged 6-9 months from site acquisition to operation, and planned operational life was 36 months to refit. Even high energy-prices were dwarfed by net daily profits of 1-2 million dollars per day for these buildings, creating little incentive for efficient use of energy.

Types of data centers

The term "data center" refers to many different buildings, all having high energy use as a commonality. The youth of the data center industry lends to a variety of terms that will solidify with continued research. We use the term "data center" to refer generally to data storage facilities, and specifically to any of the following categories:

1. Telecommunication switches ("Telecoms")
2. Internet service provider routers ("ISPs")
3. Co-located server hosting facilities ("CoLos"). These have been defined as locations that combine server hosting area and office space or other working space. According to a Solomon Smith Barney (2000) report, collocation facilities are the fastest growing types of data centers. These locations are difficult targets for energy-efficient retrofits because they are not specifically designed to house servers, and therefore may have complicated power needs.
4. Data storage and hosting facilities ("server farms" or "internet hotels"). These facilities are built specifically for data storage, and often are maintained by a single company (even if it is a company that rents out servers to outsourcing groups), and therefore the whole building can be built or retrofitted to the owners needs, including energy needs.
5. Corporate datacenters, including both servers and mainframe computers. These are the oldest types of data centers.

In addition, corporate "call centers," where company staff respond to technical support, customer service, telemarketing and reservations calls are sometimes included in data center discussions. The power densities in call centers are lower because of the lower fraction of the facility space dedicated to computer equipment.

Early reports regarding the energy use in data centers were mostly from within the industry in the form of newspaper, investment or trade articles (Bryce 2000, Hall 2001, Joyce 2001, etc). More recently, however, reports have started to surface from surrounding industries (i.e. construction and energy industries) affected and influenced by the Internet industry. Also, because of the public goods nature of the internet, the government has performed some studies regarding the internet backbone that have been beneficial to understanding the physical workings of the internet, and the role of high tech buildings (FCC 2000). Most recently, reports related to energy use and efficiency in high tech buildings have begun to surface (Mitchell-Jones 2001, Laitner et al 2000). Although the volume of information is increasing, future projections for installation and even current estimations of installed base are still widely varied.

Current Installations

Estimates of total installed base are uncertain. Based on requests for service of utility companies, year 2000 projections for the 2000-2002 time frame were 400-800 MW at each node of the fiber backbone. The 35 primary nodes (as noted in Figure 2) would create approximately 15 GW total capacity, representing more than twice the estimate of likely installed capacity made by Jon Koomey at LBNL (2000). The actual number may be closer to Koomey's estimate, given that only a fraction of the requests for service result in actual installations. That more power is requested than is used is a problem with knowledge about the true requirements, and is also indicative of the operational mentality of the data center industry: having enough power is paramount. The economics of the industry depend greatly on adequate and reliable power, as even milliseconds of downtime can cost companies millions of dollars. A representative of Exodus Communications indicated at a data center design charette in February of 2001, outages are not acceptable in data centers. In addition, some data center developers have used the power capacity as a proxy (all be it improperly) for reliability of power supply.

In the current market, utilities are attempting to lessen the amount of overestimation on the part of data center builders. Before the market slowdown, utilities could expect to recoup the cost of building enough infrastructure to satisfy the requests for power. Now, however, many utilities have installed "use it or pay for it" programs to make sure they do not waste money on infrastructure that will not be used.

Future Projections

Another problem is determining the accurate projection of future growth. This is made more difficult by rapid rise of these facilities since 1998 (Figure 1). The change in market dynamic late in the year 2000 will show a slower growth rate in data centers, which may change the way they are built, and allow energy-efficient technologies to be incorporated into design.

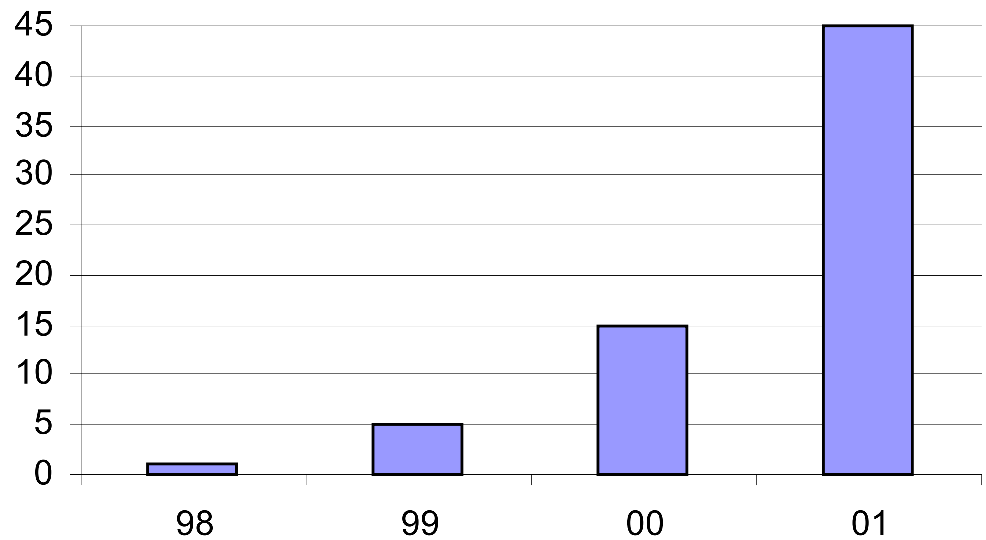

Figure 1: Estimated capacity of data centers worldwide. Millions of gross square feet (Source: Soloman Smith Barney 2000).

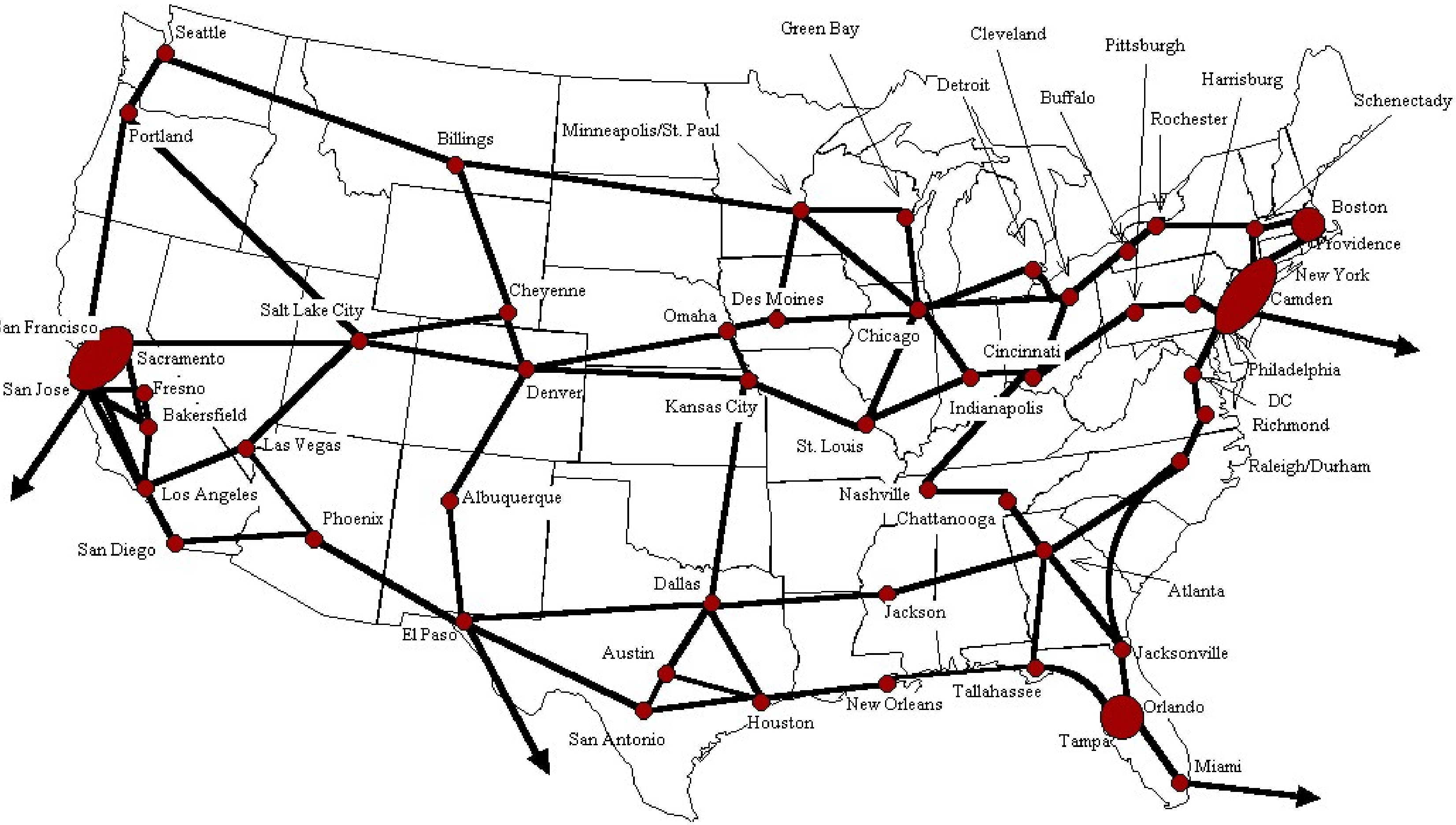

Figure 2. Map of the Internet Backbone Showing Primary Nodes (Source: FCC 2000)

The future of these high-energy-use, high-reliability buildings is a matter of significant debate. Some project that the market will saturate in the near future resulting in a steady load due to increases in throughput which are expected to occur at the rate of 10 times every 18 months. The current slow down appears to result from of over-capacity resulting from the downturn in the demand by e-commerce companies combined with some degree of over-building. However, though the health of Internet businesses has declined, we have seen many traditional businesses establish an Internet presence. The overwhelming belief in the construction and internet industries is not that the market will collapse, but that it will continue to grow, presumably at a more rational pace (Joyce 2001).

Others factors can impact the future of data center energy use. New technologies in this field may change energy needs. For instance, that optical switches and servers could reduce the total load.

Another important question is whether this is new load, or a shift of distributed servers to aggregated server farms concentrated in a few locations around the country. Some anecdotal evidence indicates the latter. A number of major server manufacturers have reduced their sales of servers to the end-users, and shifted to either providing data center services directly (e.g., IBM, Dell, Compaq) and/or selling to server hosting companies (e.g., Digex, Oracle)(Cope 2001). Despite the projections (Solomon Smith Barney 2000), this outsourcing could indicate a decrease in colocation facilities and an increase in data storage and hosting facilities, opening up the market for clean and efficient technologies.

MEETING THE DATA CENTER POWER NEED

Power in data centers is divided between powering the servers and the much larger need for quality HVAC. Today's servers generate heat by switched mode power supplies, which are very powerful and emit a substantial amount of heat. HVAC has been projected to range from 40-60% of the electric load in data centers. HVAC is estimated to be responsible for 40-60% of power use in data centers (Leake 2001).

Historically, centrifugal and absorption chillers are generally not used for cooling because of the 180-240 day ordering lead-time. In the attractive market of the late 1990's, the rush to build data centers led to the use of direct expansion (DX) packaged cooling in these facilities because of the short lead-time (30-60 days) required by these units. In addition, it has been reported that some developers are concerned about having circulating cooling water in the building.

These facilities operate 24/7 and require 99.9999% reliable power at a 50 ms level. They aim for 99% reliability from the grid, 99.99% from the addition of standby diesel generation sets, and 99.9999% by adding UPSs (uninterruptible power sources) on each server. Diesel fuel is used because:

1. Diesel generators can be delivered in less than 60 days (in the fall of 2000, Caterpillar was producing 40MW of new generators a day, which was close to full capacity, to meet this market need (Donahue 2000).) It can take over 180 days to get natural gas service to a site.
2. They have black start capability (the capability to operate at full load within 30 seconds).

To our knowledge, no problems have been encountered with obtaining environmental permits for the diesel generators since they are permitted as emergency/standby units. Many states enforce per-year limits of typically 100 to 200 hours on the operation of diesel backup generators. It is important to recognize however that diesel generators emit nearly 20 times the amount of nitrous oxides of clean natural-gas engines (Shipley et al. 2001). This disparity in emissions levels becomes even greater when a natural-gas engine is operated in CHP mode.

It is easy in most states to install standby generators as they are generally used for emergency power to hospitals and other facilities that require constant power. Most state and local air quality authorities exercise limited oversight of this equipment. Several state air quality agencies and environmental organizations, including state agencies in Texas and California, have expressed concern about data centers for several reasons:

1. There has been a very large increase in the installed base of emergency generators.
2. There is a concern/expectation that these units are likely to run many more hours than generators in conventional applications (e.g., hospitals have typically operated less than 40 hours per year, while reliability power can operate several hundred hours per year).
3. These diesel generators are more likely to run during peak power periods, which correspond with air quality non-attainment events. Conventional emergency generation operates during grid outages that are most frequently caused by severe weather.

These concerns reflect that these units may result in significant air quality impacts that are not modeled, and that existing regulation may not be structured to regulate these sources adequately. Assuming that these standby diesel generators are operated on 100 hours per year (the limit in many states), and that they are uncontrolled, a 40 MW site would emit 24 tons of NO_x, nearly single-handedly triggering the threshold for new source review by the Environmental Protection Agency. 40 MW of modern turbines without selective catalytic reduction (SCR) technology could operate all year at full load and emit less. However, the relaxed nature of standby generation regulation and more stringent distributed generation permitting perversely encourage the installation of diesels over more efficient and less environmentally damaging technologies.

CHP POTENTIAL AT DATA CENTERS

Combined heat and power (CHP) technologies may offer a reliable, energy efficient alternative for providing data centers with high-quality power. Combined heat and power systems are highly efficient, reliable and offer flexibility in fuel selection. Modeling analysis has demonstrated significant air emissions, transmission and price benefits of clean CHP technologies (Center for Clean Air Policy 2000). Despite these benefits, CHP remains an underutilized technology hindered by a number of disincentives. These barriers can be summarized as follows (Elliott and Spurr 1999):

- Complicated permitting systems that are complex, time consuming, and varied.
- Current regulations do not account accurately for the overall system efficiency of CHP, or credit displaced emissions and grid losses.
- Difficult and frequently prohibitive interconnection arrangements with utilities.
- Depreciation schedules that do not reflect the true life of CHP assets.

One of the greatest barriers to installation of CHP is the complicated and lengthy plant siting and permitting process. In non-attainment areas, major new sources are required to meet New Source Review (NSR) requirements to obtain operating and construction permits. NSR sets stringent emission rates for criteria pollutants and requires the installation of the best available control technology. New sources are also required to offset existing emissions in non-attainment areas. Nevertheless, current emissions standards are generally based on fuel input, an approach that does not recognize the fuel efficiency CHP technologies. Moreover, non-uniform interconnection standards and unfair utility tariffs inhibit the installation of CHP and other distributed energy resources.

Currently, there bills being introduced in Congress to ease this disparity and create a better market for CHP. The timing of the legislation falls well for the current market for data centers. As the huge profits of the year 2000 begin to soften, data center builders and owners look more carefully at how to lower the costs of these facilities in order to increase profits. Lowering costs means streamlining the power needs, and even slowing down the time to market so as to prepare for a more long-term data storage facility. It is in this market atmosphere that clean and reliable technology will make the largest impact.

RECOMMENDATIONS

Careful steps and preparation for increased load due to data center growth can also prevent possible power crises like the one in California in the winter of 2000. Options to meet these new loads in an efficient and environmental beneficial manner are available, including efficient CHP systems providing both cooling and power. These systems address the inability of the grid to respond to these new, concentrated loads, while providing greater reliability with less environmental impact relative the current grid supplied power with diesel backup.

Also, the efficiency opportunities inside these buildings are tremendous. Little or no effort is expended in optimizing the energy use of the building systems, but with profits falling, these cost-saving measures may become more popular. More efficient HVAC could significantly reduce electricity usage, decreasing load and cost of the facilities. Alternatively, a shift to direct cooling of the equipment could improve efficiency and allow greater equipment density than is allowed with existing convection cooling. This is of particular interest to an industry that generally holds to Moore's law. Moore's law is expected to hold until 2005 (Parson 2000), and states that semiconductor performance will double every 18 months. This means power densities will continue to increase, increasing the need for high-powered HVAC systems. In addition, these building systems are likely to last several generations of servers, as we have seen with the semiconductor industry, so the investments made today will continue to show returns in both energy and cost savings for years to come.

Opportunities also exist with the equipment installed in the buildings. Much of the inefficiency and thermal load results from the power supplies, not from the servers and routers. Moving to redundant DC power distribution in the building could remove a large thermal load from the conditioned space while improving reliability and availability.

There is a great deal of opportunity for data center energy efficiency to be increased, lowering the cost of maintenance, and maintaining the required reliability demanded by the industry, avoiding the use of polluting diesel generators. It is possible that the current market slowdown will allow for more planning and building time for these facilities, thus allowing for better use of reliable and clean power.

REFERENCES:

Cope, James. 2001. "Update: MetLife Outsources Network Management to AT&T." *ComputerWorld* [online]. http://www.computerworld.com/cwi/story/0,1199,NAV47_STO55072,00.html.

Center for Clean Air Policy. 2001. *Promoting Clean Power, Clean Air and Brownfield Redevelopment*. Washington, DC: Center for Clean Air Policy.

Donahue, John. 2000. Personal communication to Neal Elliott, October. Lafayette, Indiana: Caterpillar Inc.

Elliott, Neal and Mark Spurr. 1999. *Combined Heat and Power: Capturing Wasted Energy*. Washington D.C.: American Council for an Energy-Efficient Economy.

[FCC] Federal Communication Commission. 2000. *Deployment of Advanced Telecommunications Capability*. Washington, DC: FCC.

Hall, Mark. 2001. "Net Blamed as Crisis Roils California". *Computer World* [online]. http://www.computerworld.com/cwi/story/0,1199,NAV47_STO56341,00.html

Joyce, Amy. 2001. “Big Data Centers Stand Empty.” *Washington Post Online.* http://washingtonpost.com/wp-dyn/articles/A62991-2001May22.html.

Koomey, J. 2000. *Re-estimating the Annual Energy Outlook 2000 Forcast Using Updated Assumptions about the Internet Economy. Presented at the Eastern Economics Conference*. Available at http://enduse.lbl.gov/proects/infotech.pdf. Crystal City, Va. March 24.

Leake, Ernie. 2001. Personal communication at *Austin Energy Data Center Design Charette*, February 12-13, 1001, Austin, Texas. Exodus Communication.

Mitchell-Jackson, Jennifer. 2001. *Energy Needs in an Internet Economy: A Closer Look at Data Centers*. Masters thesis. Under Review. Berkeley, Calif.: University of California.

Mullen, Theo. 2000. "E-Commerce Short On Juice." *Internet Week* [online]. http://www.internetwk.com/lead/lead062600.htm

Parson, Ellen. 2000. “Tapping into the Data Center Market.” *Industry Click* [online]. http://www.industryclick.com.

Rohde, Laura. 2001. "Verizon, Sprint Unveil Global Network Plans." *ComputerWorld* [online]. http://www.computerworld.com/cwi/story/0,1199,NAV47_STO57479,00.html

[Salomon Smith Barney] Mahedy, Stephen, D. Cummins and D. Joe. 2000. "Internet Data Centers: If Built…Will They Come?" New York, NY: Salomon Smith Barney

Shipley, Anna, Neal Elliott, Nathanael Greene, Jia Li, and Katie McCormack. 2001. (Publication Pending). *Certification of CHP Systems*. Washington D.C.: American Council for an Energy-Efficient Economy.

Wagner, Mitch. 2001. "California E-Businesses Weather Power Outages." *Internet Week*[online]. http://www.internetweek.com/story/INW20010119S0005.

Walker, Megan. 2000 . "Silicon Valley Braces For More Energy Shortages." *Techweb/Internet Week* [online]. http://www.internetweek.com/story/INW20001228S0003