Data Centers

Information technology continues to predominate in the twenty-first century, and the data centers that house the increasing numbers of computers and servers needed to meet demand may strain energy sources. Not only is adequate and continuous power to supply technology an issue, but the growing demand for water to cool equipment may become a concern as well.

If you use Facebook, MySpace, YouTube, iTunes, Google, online banking, online gaming, an online weather service, an email program, or if you read and/or create blogs, all your activities are housed in a data center. Data centers are rooms full of computers (or servers) arranged in rows that run the software and applications that allow you to do everything you do on the Internet. Small data centers have only a few servers, but the ones that support so much of what we do on the Web can house thousands, and in some cases, hundreds of thousands, of servers. For example, some estimates put the number of servers in use by Google at over 400,000.

Imagine a building in which the floor is lined with rows of metal racks. Each rack contains a stack of servers. All of these servers run twenty-four hours a day, seven days a week, and while they run, they constantly create a tremendous amount of heat. (Put your running laptop on your lap for a few minutes and you'll get the idea.) To prevent the room from overheating to a temperature that affects the reliability of the servers, the air must be kept cool. So in addition to rows of servers, data centers require infrastructure, which consists of equipment that supplies the power (uninterruptible power supply) and the cooling equipment that keeps the data center's air from getting too hot.

Data centers also contain networking and storage appliances that hold and transmit information within a company, between companies, or to consumers. Often the rows of servers and appliances are in the middle of a data center, and the infrastructure lines the walls.

Data centers are measured in various ways. Until 2002 (or thereabouts), they were measured in units of area. A 100-square-meter data center can run a small law firm, accounting firm, real estate company, or the like. Today data centers are often described in terms of their power requirements—how many kilowatts they require to power the infrastructure and the information technology (IT) equipment itself.

Types of Data Centers

There are four different types of data centers. The characteristics of each serve the needs of different business models.

Internet Server Farms

Corporations like Amazon, Google, Microsoft, and Yahoo house this type of data center. Because Internet server farms are very large and accommodate great numbers of servers, they must be well planned and built in locations that provide special conditions, including access to a plentiful supply of power and attractive utility rates, availability of lots of water (for cooling), proximity to Internet data lines (called fiber), and safety from the potential for terrorist threats and weather- or geography-related disasters.

Collocation Services (CoLo)

Collocation services are data centers built and run by companies like Savvis, Equinox, and Switch & Data that manage services for several companies in one or more locations. They rent out space, and sometimes equipment as well, in these data centers to other companies who put their

applications and data on the rental servers. The CoLo companies take care of running the building and tending to the servers. A collocation service might house data centers for a number of different companies in the same building.

Companies use collocation services for many reasons. Some of the chief incentives are that businesses can:

- save on the cost of owning and operating a data center
- focus on their core enterprise, not the expertise needed to operate data centers in-house
- depend on the high level of security and reliability built into collocation services
- plan for expansion space at comparatively reasonable rates
- add data center capacity quickly

Enterprise Data Centers

An enterprise data center is owned and operated by the corporation that uses it. Typically, enterprise data centers are established in the office building that houses the corporation. Because this space (465–1,900 square meters on average) often preexists the data center, it may be far from optimal for data center use and often must be retrofitted to accommodate it. As the company grows and the demands on the data center expand, the company may have to tackle difficult and costly problems related to supplying extra space, power, and cooling.

Server Closets

Smaller businesses meet their data center needs by using server closets, which are small rooms (and in some cases, literally closets) less than 465 square meters in size. These spaces contain less infrastructure but still must be well-cooled to prevent heat buildup from disabling the servers. All of servers in these rooms may be shut down from time to time for maintenance; this would seldom occur in the large data centers described above.

Data Center Infrastructure

Server farms, collocation data centers, and enterprise data centers contain a complex array of infrastructure. The more reliable the data center—meaning it must continually provide its services without any interruptions—the more infrastructure it requires. Infrastructure consists of the equipment that supplies both power and cooling, the cabling that connects all of the equipment, networking equipment that allows servers to connect to each other and to the Internet, storage equipment (for data storage and retrieval), security equipment, fire prevention equipment, and more.

Over the years the computing power of servers and the number of servers used in data centers have dramatically increased. This has driven up the amount of cooling that data centers require, which in turn has dramatically increased the amount of power that data centers consume. All of this has greatly enlarged their infrastructure requirements.

As an example, provisions have to be made to keep the power flowing to the equipment in data centers that can't allow any interruptions of service, even in the event that the utility company that supplies power goes offline even for a fraction of a minute. So the infrastructure will include generators to supply emergency power, uninterruptible power supply (UPS) devices to make sure that power continues to flow, backup batteries, and the like. All of these items add to the cost and space requirements of data centers. This equipment needs to be installed, cabled, monitored, and maintained, which means dollars must be spent, not only on the IT and infrastructure equipment, but also on the people who install and maintain them as well. Every data center expansion requires more of this equipment to be installed, using up more space and more people-hours.

A direct result of this is that the space requirements of data centers have tripled since 2000. According to Tony Ulichnie, IT consultant and acting director of Site Uptime Network, Uptime Institute, ten years ago a 100-square-meter data center required an additional 30 square meters (approximately) of space to house the electrical and cooling equipment, a space roughly one-third the size of the data center. (This figure is scalable to about a 1,000-square-meter data center.) By the late 1990s, the increased density in that same 100-square-meter data center required an additional 50 square meters of space to house the electrical and cooling equipment. Currently, a 100-square-meter data center requires a room of equal size to house the electrical and cooling equipment (personal communication with author).

Given these infrastructure requirements, it would seem logical for data center managers to maximize the use of every single square meter in the facility and every kilowatt of power. But often, that's not what actually happens.

Energy Crises in Data Centers

There is a surprising underside to data centers:

> Data center power and cooling infrastructure worldwide wastes more than 60,000,000 megawatt-hours per year of electricity that does no useful work powering IT equipment. This represents an enormous financial burden on industry and is a significant public policy environmental issue. (Rasmussen 2008, 2)

(Neil Rasmussen, the Chief Innovation Officer of American Power Conversion [APC], calculated this figure based on APC's estimates of the installed operating megawattage [one megawatt equals one million watts] of global data centers, along with best estimates of how much of the energy used by those data centers was "wasted," meaning that the data center was not performing up to its ability due to issues of equipment positioning, oversizing, and improper settings and adjustments. APC further found that a typical data center was wasting around 20 percent of its total power draw in this way.)

According to analysts at McKinsey & Company, the data center industry will contribute more carbon emissions than the entire airline industry by 2020 (Haskins 2008). A report from the U.S. Environmental Protection Agency (EPA) claims that at the present rate of growth in power consumption by data centers, ten more nuclear or coal-fired power plants will be required to meet this demand in the United States by 2011 (2007, 58).

Regrettably, data centers are not models of effective management or energy-efficient processes and technology. In fact, they are contributors in two crises, one ongoing and one soon to be.

Ongoing Crisis: Power

Like most households and businesses, data centers buy power from a utility company. But data centers are very large consumers. The same EPA report cited above states that "data centers can be more than 40 times as energy intensive as conventional office buildings" (EPA 2007, 17). When a large data center operator decides to locate in a utility's territory, the utility may have to act quickly to figure out how to provide the electricity the data center requirements while still providing for the needs of all the other utility customers.

As power demand by data centers increases, the cost of operating the data center skyrockets. IT organizations are accustomed to thinking that servers are cheap; this is because they're used to an earlier time when twenty to thirty years were required for the cost of the electricity powering a server to exceed the cost of the server itself. If a server cost $2,500, the incremental addition of the cost of electricity over the life of the server (three to five years) was like a free ride. Now, it takes less than two years for the cost of the electricity to exceed the cost of the server (Brill 2008). Hence, servers are not a cheap resource at all. A $2,500 server can bring with it as much as $7,500 in electrical costs during its three-year lifetime before it is retired.

This situation has created a crisis for the data center industry for several reasons. At the current rate of demand, the supply of power will be insufficient to meet the needs of the industry as soon as 2011 (Brill 2008). And as the number of servers in use increases, and the costs associated with supplying power to them increase, the ability of data centers to operate profitably is severely challenged.

Coming Crisis: Water

Water is twenty-seven times more effective in dissipating heat than air, and thus it is commonly used in data centers. The water is either circulated through the data center in pipes or piped directly through the racks to the servers. Pipes carry the heat away from the servers to a cooling tank, where the heat evaporates.

According to Dr. Robert Sullivan, a cooling expert and consultant to the Uptime Institute, if a data center operates at an average industry efficiency, approximately 109,000 liters of water per day are required to dissipate the heat that is produced by one megawatt of IT load (the amount of power drawn by the IT equipment) (personal communication with author). To put this in context, large data centers can draw quite a bit more than one megawatt of power. For example, an Air Force data center built in 2007 was designed to draw a load of 50 megawatts (Miller 2007). Microsoft's new data center in Chicago, Illinois, will draw a load of 60 megawatts (Josefsberg 2009). If these data centers were water cooled (rather than air cooled), the amount of water required would be truly staggering (Fonteccio 2008).

Currently, most data center operators take the relatively low price and ready availability of water for granted. But according to a study by the environmental think tank Pacific Institute, "water is becoming scarcer globally and every indication is that it will become even more so in the future. Decreasing availability, declining quality, and growing demand for water are creating significant challenges to businesses and investors who have traditionally taken clean, reliable and inexpensive water for granted. These problems are already causing decreases in companies' water allotments, shifts toward full-cost water pricing, more stringent water quality regulations, growing community opposition, and increased public scrutiny of corporate water practices" (Pacific Institute 2009, 1).

Limitations in the availability of water could hit the data center industry hard. The growing difficulty of securing the required amounts of water, and the higher cost of available water, may threaten the economic viability of some data centers.

Managing Data Centers

One of the hurdles to data center efficiency is that, in an enterprise, data centers are typically managed by two separate organizations: the IT group and the facilities organization.

The IT department, usually under the chief information officer (CIO), is in charge of the servers, the applications that run on them, the networking equipment through which they communicate, and the storage equipment where the data is archived. For this group, availability (uptime) has the highest value. IT wants those servers up and running as much as possible because IT compensation incentives are often based on preventing or minimizing disruption.

The facilities organization, often under a vice president of corporate real estate, is in charge of the physical plant, the cooling apparatus, and all of the equipment involved in supplying power to the data center. Essentially, this group delivers the kilowatts and cooling that IT operations require. Typically, the power bill for the entire data center operation (including the portion used by IT) is paid for by the facilities organization; in fact, most CIOs never see the power bill. Since facilities organizations typically pay the utility bill, they have a financial incentive to try to contain costs, and because CIOs have no stake in the bill, saving energy is not a high priority for them.

Old habits and outdated operational procedures often keep the facilities and IT organizations from working together effectively. Their divergent interests and responsibilities fracture the view of the data center that company executives receive. Lacking an integrated understanding of the entire data center, company executives keep looking for the same things that led to reliability and success in the past (such as redundancy), meanwhile ignoring the reality that power costs are rising steeply, that carbon emissions are growing rapidly, and that both may threaten the cost effectiveness of their data centers if not controlled.

To remedy this situation, Will Forrest, industry analyst and principal at the management consulting firm McKinsey & Co., claims that the cooperation of these two groups is essential. In fact, he recommends that the facilities organization be placed under the control of the CIO so that the power bill and concerns of each group will be integrated (Kaplan, Forrest, and Kindler 2008).

Future of the Data Center

The data center industry is in transition for a variety of reasons. There are both short-term and long-term influences. The more immediate reasons for this transition include:

- the increasing complexity and difficulty of managing data centers
- the uncertainty of infrastructure guidelines from government and advisory organizations
- the prohibitive cost of new data center construction costs and the difficulty of financing them
- the availability of alternatives to owning a data center in the form of collocation facilities that manage the infrastructure
- improving technologies for cooling the data center

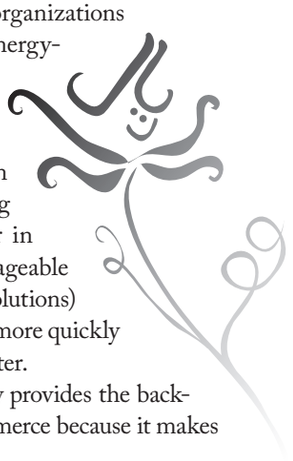
The long-term issues pushing data centers into transition include:

- more competition over the availability of water for cooling
- alternatives that are more cost competitive with different data center ownership
- semiconductor chips and operating-systems software coding that are more energy aware and efficient
- government regulation of carbon emissions using cap and trade incentives or regulation

Where are these influences leading the data center industry? Some experts predict that rather than owning data centers, some businesses will elect to buy data center services provided over the Internet. This is known as cloud computing. (Think of how Google provides all its services via the Internet.) It may be that this takes the form of "clouds" owned and operated by a provider such as Google, Amazon, Yahoo, or Microsoft.

Some experts predict a boom in the use of collocation facilities. According to these experts, as data center operations and ownership become more costly and complex, companies will want to focus more on their core business and less on their data center operations. Combined with the need for increased computing power, this makes cost-effective collocation opportunities much more attractive to these companies.

Another trend is also making itself evident. Large organizations like Sun Microsystems and Dell, as well as small startup organizations, are creating modular approaches to data center design. By creating "pods" with built-in cooling and infrastructure, these organizations can deliver much more energy-efficient and secure data centers and can make expansion much less expensive. The pods can be delivered in shipping containers (Microsoft) or in smaller, individually manageable units (Elliptical Mobile Solutions) and can be installed much more quickly than a traditional data center.



The data center industry provides the backbone for business and commerce because it makes

possible the rapid transactions that businesses and individuals need to accomplish quickly around the globe. It makes economic sense to send data files instead of FedEx packages.

But the more information we move, the more we must rely on the data centers that make this happen. In fact, the data center industry can be seen as the backbone of the economy as a whole. If it is not efficient and economically viable, it negatively affects the rest of the economy. The challenge that the industry faces then is to supply an ever-greater amount of computing power while consuming a smaller portion of the electrical power and natural resources.

Deborah Puretz GROVE and David ROSENBERG
Grove Associates LLC

See also Energy Efficiency; Energy Industries *(assorted articles)*; Facilities Management; Information and Communication Technologies (ICT); Supply Chain Management; Telecommunications Industry; Water Use and Rights

Further Reading

Brill, Kenneth G. (2008, June 16). Moore's Law economic meltdown. Retrieved August 28, 2009, from http://www.forbes.com/2008/06/16/cio-moores-law-tech-cio-cx_kb_0616moore.html

CB Richard Ellis. (2008) *Viewpoint European data centres Q4 08*. Retrieved August 28, 2009, from http://www.cbre.co.uk/uk_en/services/global_corporate_services/data_centres_networks/research

Fonteccio, Mark. (2008, November 24). Experian drills backup water supply to cool data center. Retrieved August 28, 2009, from http://searchdatacenter.techtarget.com/news/article/0,289142,sid80_gci1339996,00.html

Haskins, Walaika. (2008, May 1). Data centers may spew more carbon than airlines by 2020. Retrieved September 13, 2009, from http://www.crmbuyer.com/story/62840.html?wlc=1252880067

Josefsberg, Arne. (2009, June 29). Microsoft brings two more mega data centers online. Retrieved July 25, 2009, from http://blogs.technet.com/msdatacenters/archive/2009/06/29/microsoft-brings-two-more-mega-data-centers-online-in-july.aspx

Kaplan, James M.; Forrest, William; & Kindler, Noah. (2008, July). *Revolutionizing data center energy efficiency.* Retrieved August 28, 2009, from http://www.mckinsey.com/clientservice/bto/pointofview/pdf/Revolutionizing_Data_Center_Efficiency.pdf

Miller, Rich. (2007, November 9). 50 megawatt data center in Sacramento. Retrieved August 28, 2009, from http://www.datacenterknowledge.com/archives/2007/11/09/50-megawatt-data-center-in-sacramento/

Pacific Institute. (2009, February). *Water scarcity and climate change: Growing risks for businesses and investors.* Retrieved August 28, 2009, from http://www.ceres.org/Document.Doc?id=406

Rasmussen, Neil. (2008). *An improved architecture for high-efficiency, high-density data centers* (APC White Paper #126). Retrieved August 28, 2009, from http://www.apcmedia.com/salestools/NRAN-6V5QAA_R0_EN.pdf

U.S. Environmental Protection Agency (EPA). (2007, August 2). Report to Congress on server and data center energy efficiency public law 109–431. Retrieved August 28, 2009, from http://www.energystar.gov/ia/partners/prod_development/downloads/EPA_Datacenter_Report_Congress_Final1.pdf