Reducing Data Center Energy Consumption

By John Judge, Member ASHRAE; **Jack Pouchet, Anand Ekbote,** and **Sachin Dixit**

Rising data center energy consumption and increasing energy costs have combined to elevate the importance of reducing data center energy consumption as a strategy to reduce costs, manage capacity, and promote environmental responsibility. Within almost every organization, data center energy consumption has been driven by demand for greater computing capacity and increased IT centralization. The demand has been increasing by approximately 12% per year.1 While this was occurring, U.S. electricity prices have increased by 4.4% per year.2 The financial implications are significant. Estimates of annual power costs for U.S. data centers now range as high as $3.3 billion.

The good news is that there is general agreement within the industry that improvements in data center efficiency are possible. A 2007 EPA report to the U.S. Congress concluded that best practices can reduce data center energy consumption by 50% by 2011.1 It included a list of Top 10 Energy Saving Best Practices as identified by Lawrence Berkeley National Laboratory. Other organizations have distributed similar information, and there is evidence that some of these best practices are being adopted.

While progress has been made, what has been lacking is a holistic, system-level approach useful in guiding data center managers in prioritizing opportunities to reduce energy consumption. This has made it difficult for data center managers to prioritize efficiency efforts and tailor best practices to data center equipment and operating practices.

Data Center Energy Consumption

The first step in prioritizing energy saving opportunities is to gain a solid understanding of data center energy consumption.

A recent analysis modeled energy consumption for a typical 5,000 ft2 (465 m2 data center based on real-world technologies and operating parameters and analyzed how energy is used within the facility. Energy use was categorized as either demand side or supply side. The total energy consumption of the modeled data center was 1,127 kW.

In this analysis, demand-side systems—which include processors, server power supplies, other server components, storage and communication equipment—account for 52% of total consumption. Supply-side systems include the UPS, power distribution, cooling, lighting, and building switchgear, and account for 48% of consumption. The supply-side equipment is not an independent consumer of power; its power consumption depends on the power demand.

Note that all data centers are different and savings potential will vary by facility. However, at minimum, this analysis provides an order-of-magnitude comparison for data center energy reduction strategies.

The distinction between demand and supply power consumption is valuable because reductions in demand-side energy use cascade through the supply side. For example, in the 5,000 ft2 (465 m2) data center model, (referred to as "the data center model" in the rest of the article.) a 1 W reduction at the server component level (processor, memory, hard disk, etc.) results in an additional 1.84 W savings in the power supply, power distribution system, UPS system, cooling system and building entrance switchgear, and medium voltage transformer with no further action (*Figure 1*).

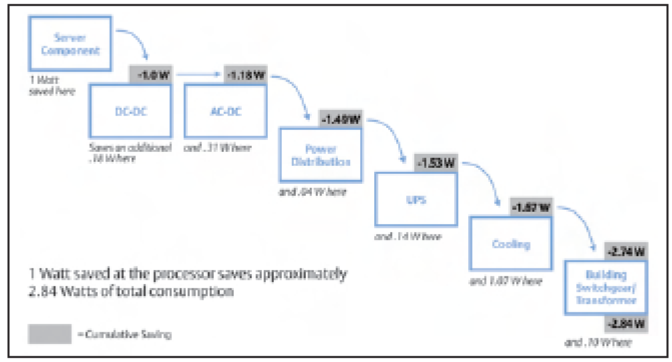

Figure 1: With the cascade effect, a 1 W savings at the server component level creates a reduction in facility energy consumption of 2.84 W.

Consequently, every watt of savings that can be achieved on the processor level creates a total of 2.84 watts of savings for the facility.

The Cascade Effect

Recognition of the energy cascade concept leads to an approach to reduce data center energy consumption that starts at the core of the data center—the server—and follows the flow of energy systematically to each subsequent level of equipment. This enables compounding benefits throughout the data center. This Energy Logic approach also removes the three most critical constraints faced by data center managers today: power, cooling, and space.

Essentially, this sequential approach is the lever that enables data center stakeholders to raise data centers to new levels of performance. It can deliver a 50% or greater reduction in data center energy consumption without compromising performance or availability. When 10 top energy-saving strategies are applied, in the order in which they have the greatest impact, to the data center model, the strategies make available two-thirds of floor space, one-third of UPS capacity, and 40% of precision cooling capacity.

While the sequence is important, it is not necessary that each step can only be undertaken after the previous one is complete. The energy-saving measures included in this sequential approach to reducing energy consumption should be considered a guide. Many organizations will already have undertaken some measures at the end of the sequence or will have to deploy some technologies out of sequence to remove existing constraints to growth.

10 Energy-Saving Strategies

1. Processor power. In the absence of a true standard measure of processor efficiency comparable to the fuel efficiency standard for automobiles, thermal design power (TDP) serves as a proxy for server power consumption.

The typical TDP of processors in use today is between 80 and 103 W (91 W average). For a price premium, processor manufacturers provide lower voltage versions of their processors that consume, on average, 30 W less than standard processors. Independent research studies show these lower power processors deliver the same compute performance as higher power models.4

In the data center model mentioned earlier, low power processors create a 10% reduction in overall data center power consumption.

2. Power supplies. As with processors, many of the server power supplies in use today are operating at efficiencies below what is currently available. The U.S. EPA estimated the average efficiency of installed server power supplies at 72% in 2005.1 In the data center model, it is assumed the unoptimized data center uses power supplies that average 79% across a mix of servers that range from four years old to new.

Best-in-class power supplies are available today that deliver efficiency of 90%. Use of these power supplies reduces power draw within the data center by 124 kW or 11% of the 1,127 kW total. As with other data center systems, server power supply efficiency varies depending on load. Some power supplies perform better at partial loads than others and this is particularly important in dual-corded devices where power supply utilization can average less than 30%. *Figure 2* shows power supply efficiencies at different loads for two power supply models. At 20% load, Model A has an efficiency of approximately 88%, while Model B has an efficiency closer to 82%.

Another opportunity to reduce energy consumption is to size power supplies closer to actual server load. Server manu-facturers size the power supply for the maximum possible server configuration, while the servers may be shipped out with considerably less power requirement. Most servers do not get upgraded to the maximum configuration; so these power supplies operate at small part loads, at low efficiency levels. End users should ask server manufacturers to size their power supplies for the configuration they intend to use.

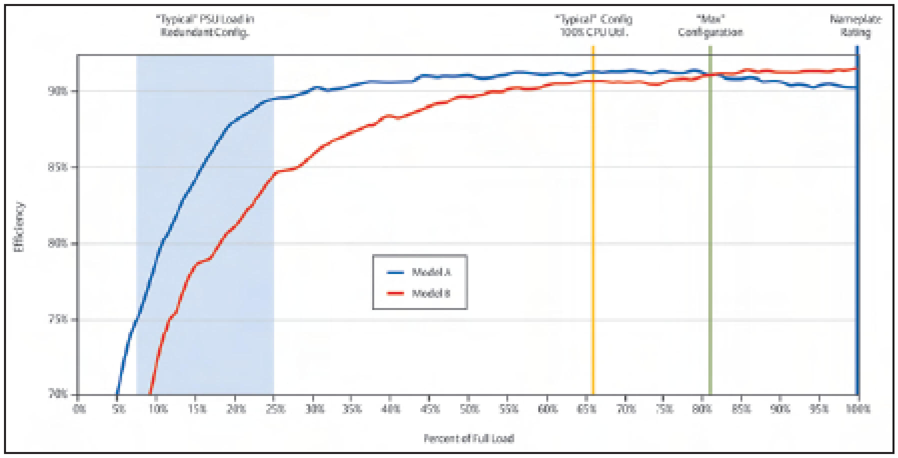

Figure 2: Power supply efficiency can vary significantly depending on load, and power supplies are often sized for a load that exceeds the maximum server configuration.

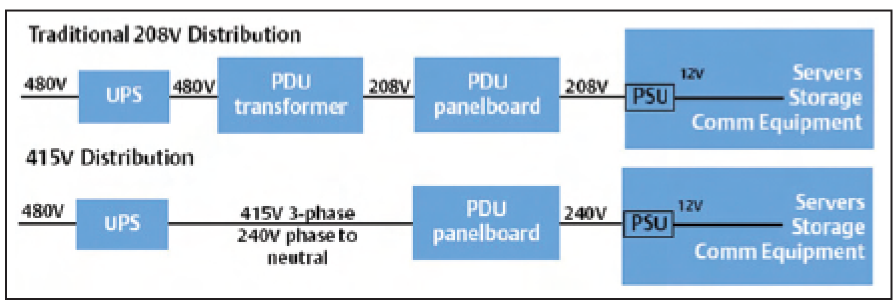

Figure 3: 415V power distribution provides a more efficient alternative to using 208V power.

3. Power management software. Data centers are sized for peak conditions that may rarely exist. In a typical business data center, daily demand progressively increases from about 5 a.m. to 11 a.m. and then begins to drop again at 5 p.m.

Server power consumption remains relatively high as server load decreases. In idle mode, most servers consume between 70% and 85% of full operational power. Consequently, a facility operating at just 20% capacity may use 80% of the energy as the same facility operating at 100% capacity.

Server processors have power management features built-in that can reduce power when the processor is idle. Too often these features are disabled because of concerns regarding response time. However, this decision may need reevaluation in light of the significant savings this technology can enable.

In the data center model, it is assumed that idle power draw is 80% of the peak power draw without power management, and reduces to 45% of peak power draw as power management is enabled. With this scenario, power management can save an additional 86 kW or 8% of the unoptimized data center load.

4. Blade servers. Many organizations have implemented blade servers to meet processing requirements and improve server management. While the move to blade servers is typically not driven by energy considerations, they can play a role in energy consumption.

Blade servers consume about 10% less power than equivalent rack mount servers because multiple servers share common power supplies, cooling fans, and other components.

In the data center model, there is a 1% reduction in total energy consumption when 20% of rack-based servers are replaced with blade servers. More importantly, blades facilitate the move to a high-density data center architecture, which can significantly reduce energy consumption. (*See Strategy 9 on High Density Supplemental Cooling.*)

5. Server virtualization. As server technologies are optimized, virtualization is increasingly being deployed to increase server utilization and reduce the number of servers required.

In the data center model, it is assumed that 25% of servers are virtualized with eight nonvirtualized physical servers being replaced by one virtualized physical server. We also assume that the applications being virtualized were residing in single-processor and two-processor servers and the virtualized applications are hosted on servers with at least two processors.

Implementing virtualization provides an incremental 8% reduction in total data center power draw for the 5,000 ft2 (465 m2) facility used as the data center model.

6. Cooling best practices. Most data centers have implemented some best practices, such as the hot-aisle/cold-aisle rack arrangement. Potential exists in sealing gaps in floors, using blanking panels in open spaces in racks, and avoiding mixing of hot and cold air. ASHRAE has published several excellent publications on these best practices, including *Datacom Equipment Power Trends and Cooling Applications*.

Computational fluid dynamics (CFD) can be used to identify inefficiencies and optimize data center airflow. Many organizations offer CFD imaging as part of data center assessment services focused on improving cooling efficiency.

Additionally, temperatures in the cold aisle may be able to be raised if current temperatures are below 68°F (20°C). Chilled water temperatures can often be raised from 45°F to 50°F (7°C to 10°C).

In the data center model, cooling system efficiency is improved 5% simply by implementing best practices. This reduces overall facility energy costs by 1% with virtually no investment in new technology.

7. 415V ac power distribution. The critical power system represents another opportunity to reduce energy consumption. However, even more than with other systems, care must be taken to ensure reductions in energy consumption are not achieved at the cost of reduced equipment availability.

Most data centers use a type of UPS called a double-conversion system. These systems convert incoming power to dc and then back to ac within the UPS. This enables the UPS to generate a clean, consistent waveform for IT equipment and effectively isolates IT equipment from the power source. UPS systems that don't convert the incoming power—line interactive or passive standby systems—can operate at higher efficiencies because of the losses associated with the conversion process. These systems may compromise equipment protection because they do not fully condition incoming power.

A bigger opportunity exists downstream from the UPS. In most data centers, the UPS provides power at 480V, which is then stepped down via a transformer, with accompanying losses, to 208V in the power distribution system. These stepdown losses can be eliminated by converting UPS output power to 415V.

The 415V three-phase input provides 240V single-phase, line-to-neutral input directly to the server (*Figure 3*). This higher voltage eliminates stepdown losses and enables an increase in server power supply efficiency. In the model, an incremental 2% reduction in facility energy use is achieved by using 415V ac power distribution.

8. Variable capacity cooling. Data center systems are sized to handle peak loads, which rarely exist. Operating efficiency at full load is often not a good indication of actual operating efficiency. Newer technologies, such as digital scroll compressors and variable frequency drives in computer room air conditioners (CRACs), allow high efficiencies to be maintained at partial loads.

Digital scroll compressors allow the capacity of room air conditioners to be matched exactly to room conditions without turning compressors on and off. This minimizes over-cooling and reduces cyclic losses in the refrigeration cycle.

Typically, CRAC fans run at a constant speed and deliver a constant volume of airflow. Converting these fans to variable frequency drive fans allows fan speed and power draw to be reduced as load decreases. Fan power is directly proportional to the cube of fan rpm and a 20% reduction in fan speed provides almost 50% savings in fan power consumption. These drives are available in retrofit kits that make it easy to upgrade existing CRACs with a payback of less than one year.

In the chilled water-based air-conditioning system used in this analysis, the use of variable frequency drives provides an incremental saving of 4% in data center power consumption.

9. High density supplemental cooling. Traditional room-cooling systems have proven very effective at maintaining a safe, controlled environment for IT equipment. However, optimizing data center energy efficiency requires moving from traditional data center densities (2 kW to 3 kW per rack) to an environment that can support much higher densities (in excess of 30 kW).

This requires implementing an approach to cooling that shifts some of the cooling load from traditional CRAC units to supplemental cooling units, which are mounted above or alongside equipment racks, and pull hot air directly from the hot aisle and deliver cold air to the cold aisle. It's important to note that the supplemental cooling units that improve data center efficiency are not spot coolers that reject more heat into the datacenter, but rather those that carry the heat out of the data center.

As compared to conventional CRACs, supplemental cooling units can reduce cooling costs by up to 30%. These savings are achieved because supplemental cooling brings cooling closer to the source of heat, reducing the fan power required to move ai. It also uses more efficient heat exchangers and delivers only sensible cooling, which is ideal for the dry heat generated by electronic equipment.

Energy Saving Action	Savings Independent of Other Actions		Energy Logic Savings with the Cascade Effect			ROI
	Savings (kW)	Savings (%)	Savings (kW)	Savings (%)	Cumulative Savings (kW)	
Lower power processors	111	10%	111	10%	111	12 to18 mo.
High-efficiency power supplies	141	12%	124	11%	235	5 to 7 mo.
Power management features	125	11%	86	8%	321	Immediate
Blade servers	8	1%	7	1%	328	TCO reduced 38%*
Server virtualization	156	14%	86	8%	414	TCO reduced 63%**
415V AC power distribution	34	3%	20	2%	434	2 to 3 mo.
Cooling best practices	24	2%	15	1%	449	4 to 6 mo.
Variable capacity cooling: variable speed fan drives	79	7%	49	4%	498	4 to 10 mo.
Supplemental cooling	200	18%	72	6%	570	10 to 12 mo.
Monitoring and optimization: Cooling units work as a team	25	2%	15	1%	585	3 to 6 mo.

* Source for blade impact on TCO: IDC ** Source for virtualization impact on TCO: VMware
©2007 Emerson Network Power

Figure 4: Using the model of a 5,000 ft^2 data center consuming 1,127 kW of power, the actions included in the approach work together to produce a 585 kW reduction in energy use.

In the data center model, 20 racks at 12 kW density per rack use high density 23supplemental cooling while the remaining 40 racks (at 3.2 kW density) are supported by the traditional room cooling system. This creates an incremental 6% reduction in overall data center energy costs. As the facility evolves and more racks move to high density, the savings will increase.

10. Monitoring and optimization. One of the consequences of rising equipment densities has been increased diversity within the data center. Rack densities are rarely uniform across a facility and this can create cooling inefficiencies if monitoring and optimization is not implemented. Room cooling units on one side of a facility may be humidifying the environment based on local conditions while units on the opposite side of the facility are dehumidifying.

Cooling control systems can monitor conditions across the data center and coordinate the activities of multiple units to prevent conflicts and increase teamwork. In the model, an incremental saving of 1% is achieved as a result of system-level monitoring and control.

Benefit of Following This Approach

This sequential approach quantifies the savings that can be achieved through each of these actions individually and as part of the sequence (Figure 4). Note that savings for supply-side systems look smaller when taken as part of the approach because those systems are now supporting a smaller load.

Employing this Energy Logic approach to the data center model reduced energy use by 52% without compromising performance or availability. In its unoptimized state, the 5,000 ft2 data center model supported a total compute load of 588 kW and total facility load of 1127 kW. Through various optimization strategies, this facility has been transformed to enable the same level of performance using significantly less power and space. Total compute load was reduced to 367 kW, while rack density was increased on average from 2.8 kW per rack to 6.1 kW per rack.
This has reduced the number of racks required to support the compute load from 210 to 60 and eliminated power, cooling and space limitations constraining growth. Total energy consumption was reduced to 542 kW and the total floor space required for IT equipment was reduced by 65%.

This sequential approach to reducing energy consumption is suitable for every type of data center; however, the sequence may be affected by facility type. Facilities operating at high utilization rates throughout a 24-hour day will want to focus initial efforts on sourcing IT equipment with low power processors and high efficiency power supplies. Facilities that experience predictable peaks in activity may achieve the greatest benefit from power management technology. *Figure 5* shows how compute load and type of operation influence priorities.

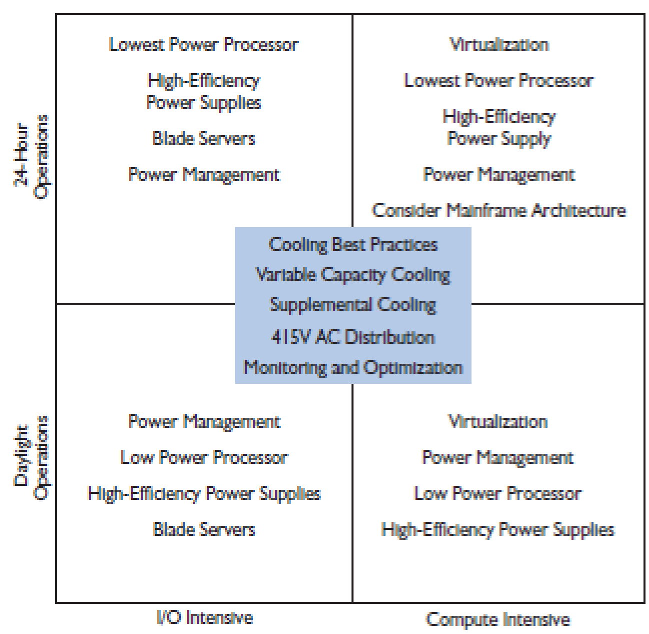

Figure 5: The sequential approach can be tailored to the compute load and type of operation.

Actions Needed to Further Efficiency

The research on this sequential approach has highlighted the need for additional work necessary to further the industry's pursuit of maximizing data center efficiency. What follows are the most significant areas requiring attention.

Define universally accepted metrics for processor, server and data center efficiency.

There needs to be an easily understood and easy-to-use measure, such as the miles-per-gallon automotive fuel efficiency ratings, which can help buyers select the ideal processor for a given load. The performance per watt metric is evolving gradually with SPEC score being used as the server performance measure, but more work is needed.

This same philosophy could be applied to the facility level. An industry standard of data center efficiency that measures performance per watt of energy used would be extremely beneficial in measuring the progress of data center optimization efforts. The PUE ratio developed by the Green Grid provides a measure of infrastructure efficiency, but not total facility efficiency. IT management needs to work with IT equipment and infrastructure manufacturers to develop the miles-per-gallon equivalent for both systems and facilities.

More sophisticated power management.

While enabling power management features provides tremendous savings, IT management often avoids this technology because the impact on availability is not clearly established. As more tools are created to manage power management features, and data is available to ensure that availability is not impacted, we should see this technology gain market acceptance. More sophisticated controls that would allow these features to be enabled only during periods of low utilization, or turned off when critical applications are being processed, would eliminate much of the resistance to using power management.

Matching power supply capacity to server configuration.

Server manufacturers tend to oversize power supplies to accommodate maximum configuration of a particular server. Some users may be willing to pay an efficiency penalty for the flexibility to more easily upgrade, but many would prefer a choice between a power supply sized for a standard configuration and one sized for maximum configuration. Server manufacturers should consider making these options available and users must be educated about the impact power supply size has on energy consumption.

Designing for high density.

A perception persists that high-density environments are more expensive than simply spreading the load over a larger space. High-density environments using blade and virtualized servers are significantly more economical on capital expenditures (reduce floor space requirement, plant capacities) and on the operating costs (energy consumption).

Depending on the details of the application, high density environments may be less likely to survive brief power outages. Mitigating this effect may necessitate adding cooling components to the UPS load.

High-voltage distribution.

415V power distribution is used commonly in Europe, but UPS systems that easily support this architecture are not readily available in the United States. Manufacturers of critical power equipment should provide the 415V output as an option on UPS systems and can do more to educate their customers regarding high-voltage power distribution.

Integrated measurement and control.

Data that can be easily collected from IT systems and the racks that support them has yet to be effectively integrated with support systems controls. This level of integration would allow IT systems, applications and support systems to be more effectively managed based on actual conditions at the IT equipment level.

Conclusion

Data center managers and designers, IT equipment manufacturers, and infrastructure providers must all collaborate to reduce data center energy consumption. In the immediacy, there are a number of actions that data center managers can take today to significantly drive down energy consumption while freeing physical space and power and cooling capacity to support growth.

By following the strategies outlined in this vendor-neutral roadmap for reducing data center energy consumption, data center managers can potentially realize a 50% or greater reduction in data center energy consumption without compromising performance or availability.

All of the technologies used in this Energy Logic approach are available today and many can be phased into the data center as part of regular technology upgrades/refreshes, minimizing capital expenditures.

References

U.S. Environmental Protection Agency. 2007. “Report to Con1. gress on Server and Data Center Energy Efficiency,” Public Law 109-431.

Energy Information Administration. 2006. “Electric Sales, Reve2. nue and Price.” www.eia.doe.gov/cneaf/electricity/esr/esr_sum.html.

Emerson Network Power. 2008. “Energy Logic: Reducing Data 3. Center Energy Consumption by Creating Savings That Cascade Across Systems,” white paper.

AnandTech. 2007. Low Power CPU Shootout.