DATA CENTERS

How data centers can minimize their energy use

Although data centers have managed to keep their collective power demand at about two percent of the world's electricity use, their energy consumption could grow exponentially if computationally intensive applications, such as video on demand, autonomous vehicles and advanced 5G technology gain wider popularity, as is expected. To manage these challenges, data centers will need to implement every possible strategy to maximize their energy efficiency.

When ABB entered the data center sector over 25 years ago, the major factors driving the market were uptime and reliability. Shortly thereafter, however, concerns about energy demand quickly stole the show. Starting in 2007, with the publication of the US EPA's seminal Report to Congress on Server and Data Center Energy Efficiency Opportunities [1], inaccuracies regarding projected data center energy use →01 resulted in the impression that the industry was becoming an insatiable energy eater. Fortunately, this impression turned out to be far from reality.

Indeed, according to the June, 2016 United States Data Center Energy Usage Report [2], combined use of electricity by all US data centers has increased only minimally in the last six years from 70 billion kWh in 2014 (representing about 1.8 percent of total U.S. electricity consumption) to an estimated 73 billion kWh in 2020 →02. What's more, this statistically minor uptick in energy demand, which has held steady at approximately two percent of overall US energy usage, has occurred in the context of a vast proliferation in the number of smart devices and the expansion of online culture to the point of near ubiquity.

Dave Sterlace
ABB Data Center Solutions
Philadelphia, PA,
United States

dave.sterlace@
us.abb.com

Key to these gains have been the exploitation of some fairly low-hanging fruits such as running data centers at higher temperatures, using virtualization to cut down on the number of underutilized servers, improved efficiency of uninterruptible power supplies (UPS), and the use of frequency drives vs dampers to control fan loads.

Other factors have also contributed to keeping data center power demand in check. For instance, servers, storage devices and infrastructures have become steadily more efficient. In addition, the industry has benefited from the trend toward larger and more efficient cloud and hyperscale centers. The latter, according to the IEA, consume proportionally much less energy for cooling compared to smaller data centers and, according to recent statistics, represent a steadily growing proportion of all data traffic. However, striving for the ultra-high efficiencies found at the state-of-the-art data centers of large web-based companies is usually not technically or economically feasible. For them every watt counts. What, then are the short-term, tactical actions that can provide immediate benefit, yielding significant electric utility cost savings?

Measuring efficiency

For years, lowering power usage effectiveness (PUE) has been high on data center operators' wish lists. Simply put, PUE is a measure of a facility's total power delivered divided by its IT equipment power usage, and all agree that the lower this figure is, the better. A PUE rating of 1.0 would be equivalent to a 100 percent efficient facility. Typically, however, data centers average about 1.67, which means that for every 1.67 watts of electricity drawn by the facility, only 1 watt is being delivered to IT equipment.

—

Over the last six years, combined use of electricity by all US data centers has increased only minimally despite vast proliferation of smart devices and online culture.

Total facility power is measured as the power dedicated to operating an entire data center. IT equipment power is defined as the power required to operate devices used to manage, route, store, or process data within a data center.

These measurements provide a baseline that allows a facility manager to compare an installation's power usage levels to that of other data centers. The only problem is that there is more than one way to calculate PUE, making it difficult to compare one facility with another.

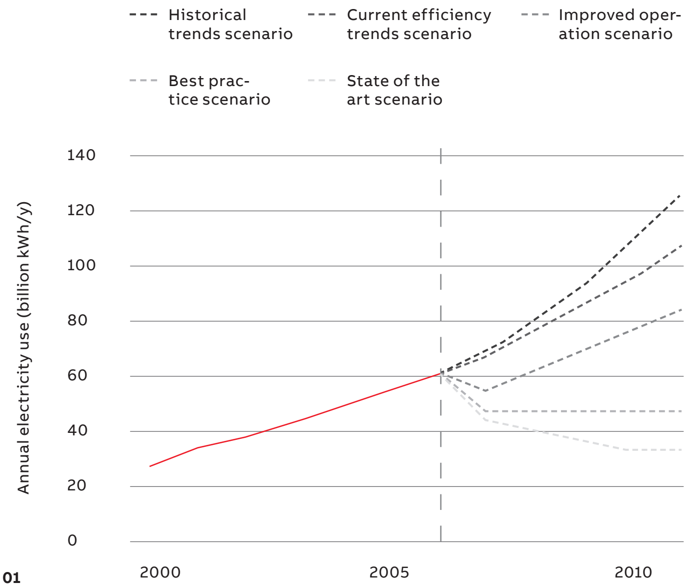
A comprehensive approach

What is needed are performance metrics that are more holistic than PUE in measuring data center efficiency. The key limitation of PUE is that it measures the overall efficiency of an entire building infrastructure supporting a given data center, indicating nothing about the efficiency of the IT equipment itself. IT efficiency, on the other hand, is the total IT output of a data center divided by the total input power to IT equipment.

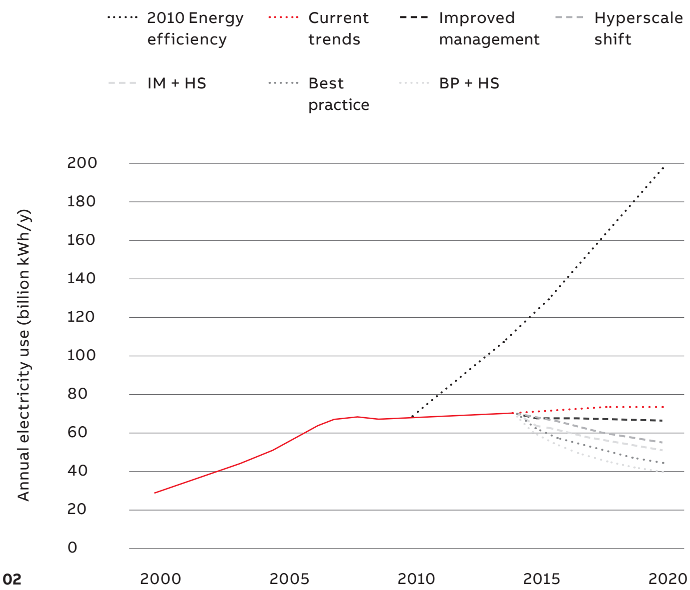
But how can IT power consumption be measured? According to the Green Grid, IT efficiency can be measured accurately after all power conversion, switching, and conditioning is complete. Thus, to correctly gauge the total power delivered to server racks, the measurement point should be at the output of the power distribution units (PDU).

Alternatively, IT output refers to the true output of the data center, in terms of the number of web pages served, or number of applications delivered. In real terms, IT output shows how efficiently the IT equipment delivers useful output for a given electrical power input. Site infrastructure efficiency indicates the amount of power that fuels actual IT equipment, and how much is diverted into support systems like back-up power and cooling. These two figures make it possible to track efficiency over time; they can also reveal opportunities as to how to maximize IT output, while lowering input power by reducing losses and inefficiencies in support systems [3].

30 percent of servers worldwide are unused. The result: $30 billion in wasted electricity per year.



01



02

Minimizing idle IT equipment

IT equipment is usually very lightly used relative to its capacity. For instance, servers tend to be only five to 15 percent utilized, processors 10 to 20 percent, storage devices 20 to 40 percent, and networking equipment is 60 to 80 percent utilized.

However, whenever such equipment is idle, it still consumes a significant portion of the power it would draw at maximum utilization. Indeed, a typical server consumes 30 to 40 percent of maximum power even when doing no work at all.

The Uptime Institute has found that 30 percent of servers worldwide are unused. This does not affect a data center's PUE but results in a loss of $30 billion in wasted electricity per year worldwide. One approach to dealing with this is distributed computing, which links computers to work as if they were a single machine. Scaling up the number of data centers that work together increases their processing power, thereby reducing or eliminating the need for separate facilities for specific applications.

03

—
01 Future electricity use by U.S. data centers as projected by the U.S. Environmental Protection Agency in 2007 [1].

—
02 Actual electricity use in all U.S. data centers, including energy used for servers, storage, network equipment, and infrastructure. Solid line represents historical estimates from 2000-2014. Dashed lines represent five scenarios through 2020 [2].

—
03 ABB SMISSLINE power distribution equipment. More than 50 percent of the power required to run a server is used by its central processing unit.

Virtualization of servers and storage
Across the industry, we can see many instances of dedicated servers and storage systems inefficiently deployed for a single application – just to maintain physical lines of demarcation. However, with virtualization, servers and storage systems can be aggregated onto a shared platform while maintaining strict segregation among operating systems, applications, data, and users.

Most applications can run on separate “virtual machines” that, behind the scenes, share hardware with other applications. Virtualization can bring great benefits for most data centers, dramatically improving hardware utilization and enabling a reduction in the number of power-consuming servers and storage devices. It can also improve server use from an average of 10 to 20 percent to at least 50 to 60 percent [4].

Consolidating servers, storage, and data centers
At the server level, blade servers can really help drive consolidation as they provide more processing output per unit of power consumed. Compared to traditional rack servers, they can perform the same work with 20 to 40 percent less energy.

Consolidating storage provides another opportunity. Since larger disk drives are more energy efficient, consolidating storage improves memory utilization while reducing power consumption.

And last but not the least, if underutilized data centers can be consolidated in one location, operators can reap vast savings by sharing cooling and back-up systems to support loads.

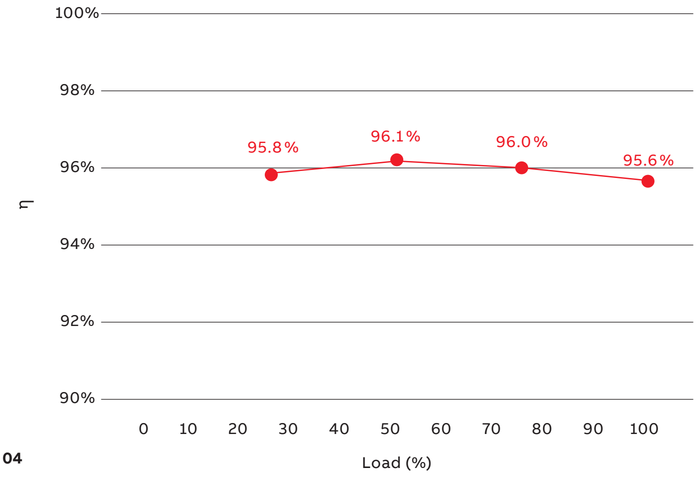
Data centers' electricity demand has remained roughly level in the past five years, in part because of a shift toward 'hyperscale' facilities, which are super-efficient due to an organized, uniform computing architecture that easily scales up to tens of thousands of servers.

On average, one server in a hyperscale center is said to be able to replace 3.75 servers in a conventional center. In a 2016 report, the Lawrence Berkeley National Laboratory estimated that energy usage would drop by a quarter if 80 percent of servers in small U.S. data centers were moved to hyperscale facilities.

Managing CPU power usage
More than 50 percent of the power required to run a server →03 is used by its central processing unit (CPU). Chip manufacturers are developing more energy-efficient chipsets, and multi-core technology enables the processing of higher loads using less power.

—
Virtualization can improve server use from an average of 10 to 20 percent to at least 50 to 60 percent.

Other options for reducing CPU power consumption are also available. Most CPUs have power-management features that optimize power consumption by dynamically switching among multiple performance states based on utilization. By dynamically ratcheting down processor voltage and frequency outside of peak performance tasks, the CPU can minimize energy waste.



04

Such adaptive power management reduces power consumption without compromising processing capability and results in significant savings when CPU utilization is variable.

Toward improved power supplies
The power supply unit (PSU), which converts incoming alternating current (AC) power to direct current (DC), consumes about 25 percent of the server's power budget, second only to the CPU. The point-of-load voltage regulators, which convert 12V DC into the various DC voltages fed to processors and multiple chipsets, is another power hog. Several industry standards, such as "Class 80+" are in place to improve the efficiency of server components.

Power drawn by a data center facility typically passes through uninterruptible power supply (UPS) and power distribution units (PDUs) before it reaches the IT equipment. PDUs generally operate at a high efficiency of 94 to 98 percent, and thus energy efficiency is primarily dictated by power-conversion in the UPS.

However, when evaluating a UPS, focusing on peak efficiency is misleading, as PDUs are unlikely to operate under a full load. Indeed, many IT systems use dual power sources for redundancy, resulting in utilization levels as low as 20 to 40 percent. In view of this, experts rely on an efficiency curve →04 to tell the whole story and to properly evaluate UPS systems. As a result, work in UPS power electronics has led to shaping the efficiency curve to be flatter and consistently higher across the range of loading.

Distributing power at higher voltages
To adhere to global standards, virtually all IT equipment is designed to work with input power voltages ranging from 100V to 240V AC. The higher the voltage, the more efficient the unit. By operating a UPS at 240/415V three-phase four-wire output power, a server can be fed directly, and an incremental two percent reduction in facility energy can be achieved [5].

Adopting best cooling practices
The cooling system of a data center contributes as much as 30 to 60 percent of its utility bill. Many facilities might have some ready opportunities to reduce cooling costs through well-established practices. But looking ahead, as server rack density keeps rising, it may be time to consider liquid cooling technologies. Traditional air-cooling systems have proven very effective at maintaining a safe, controlled environment at rack densities of two kW to three kW per rack, all the way to 25 kW per rack. But operators are now aspiring to create an environment that can support densities in excess of 30-50 kW, a level at which air-cooling systems are no longer effective [4]. In these cases, alternate cooling systems such as rear door heat exchangers may provide a solution.

On average, one server in a hyperscale center is said to be able to replace 3.75 servers in a conventional center.

Warming up to higher ambient temperature
Server rooms are usually kept at an ambient temperature of around 22 °C, leading to air conditioning unit outlet temperatures of 15 to 16 °C. However, the American Society of Heating, Refrigerating and Air Conditioning Engineers recommends temperature ranges from 15 to 32 °C for most new devices, with a humidity tolerance of eight to 80 percent.

Plugging into the smart grid
Smart grids enable two-way energy and information flows to create an automated and

—
04 A UPS efficiency curve. Experts rely on this information to evaluate UPS systems [6].

—
05 Data center operators are increasingly installing renewable power generation systems at their facilities and becoming power suppliers.

05

distributed power delivery network. They can become a key enabler for deep integration of renewable energy and distributed power generation. Furthermore, with the help of integrated monitoring and control, a smart grid can tackle the fluctuations of renewable energy, maintaining a consistent and stable power flow over the electrical grid →05.

—
A data center's cooling system contributes as much as 30 to 60 percent of its utility bill.

Data center operators can not only draw clean power from the grid, but can also install renewable power generators at a facility to become an occasional power supplier. Generators and consumers can interact in real-time, providing effective tools to receive incentive-based supply or emergency load reduction signals.

Every watt counts
From a cost and stewardship perspective, every watt counts. Taken together, all of the incremental improvements mentioned above can add up to a very significant energy-reduction impact. •

—
References

[1] US EPA Energy Star Program, "Report to Congress on Server and Data Center Energy Efficiency, Public Law 109-431," August 2, 2007. Available: https://www.energystar.gov/buildings/tools-and-resources?search=2007+Report+to+Congress&sort_bef_combine=title+ASC

[2] A. Shehabi, et al. Lawrence Berkeley National Laboratory, "United States Data Center Energy Usage Report," June, 2016. Available: https://eta-publications.lbl.gov/sites/default/files/lbnl-1005775_v2.pdf

[3] D. Cole, No Limits Software, "Data Center Energy Efficiency – Looking Beyond PUE," 2011. Available: https://www.missioncriticalmagazine.com/ext/resources/MC/Home/Files/PDFs/WP_LinkedIN%20DataCenterEnergy.pdf

[4] A. Mpitziopoulos, tom's Hardware, April 12, 2018. Available: https://www.tomshardware.com/news/what-80-plus-levels-mean,36721.html

[5] US DOE, Federal Energy Management Program, "Best Practices Guide for Energy-Efficient Data Center Design," March, 2011. Available: https://www.energy.gov/sites/prod/files/2013/10/f3/eedatacenterbestpractices.pdf