WHITEPAPER

Ten ways to drive up data center efficiency

New data centers pack higher processing power in a considerably smaller space. However, high-density computing environments can be a huge drain on operating budgets with rising energy demand and power costs. This paper discusses how to make modern data centers more energy efficient by optimizing and realizing true IT and OT efficiency.

International Energy Agency (IEA) recently estimated that although data-center workloads will triple from 2014 levels by 2020, efficiency gains mean that their electricity demand might rise by just 3 percent.[1] Even a relatively small data center can save tens of thousands of dollars simply through efficient choices in management practices, IT hardware, power, and cooling infrastructure.

Large, privately owned enterprise data center facilities still form the bedrock of corporate IT and are expected to be running half of all workloads by 2021. Striving for the ultra-high efficiencies found at state-of-the-art data centers of large web-based companies is usually not technically or economically feasible for private enterprise data centers. But, in most instances, there are short-term, tactical actions that can provide immediate benefit, yielding significant electric utility cost savings.

Measure of Efficiency

For years, a low power usage effectiveness (PUE) has been high on the wish list of any data center operator. But, PUE, a measure of a facility's total power delivered divided by its IT equipment power usage, has long been a bone of contention among professionals. However, we all agree - 'the lower the better'.

Total facility power is measured as the power dedicated to operating the whole 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 the power usage levels to other data centers. However, as per the formula provided by Green Grid, a global consortium dedicated to advancing energy efficiency in data centers and business computing ecosystems, there is more than one way to calculate PUE, making it difficult to compare one facility with another.

A PUE rating of 1.0 would be equivalent to a 100 percent efficient facility. A study by U.S. Department of Energy's Berkeley Lab found that most data centers managed a PUE in the range of 1.3 to 3.0. According to the latest survey by the Uptime Institute, an advisory body focused on IT facility certification, a typical data center has an average PUE of 1.67.[2] This means that for every 1.67 watt of electricity drawn by the facility, only 1 watt is being delivered to IT equipment.

A comprehensive approach

Due to its dynamic nature, PUE should be measured as a range of values, with their average calculated over a period. The value at any point of time depends on the number of active servers, load on the HVAC (Heating, ventilation, and air conditioning) system etc. The quality and age of the electrical and mechanical systems also plays a role. The best PUE number is often attained when all the servers are running near full capacity and shutting down redundant servers to save power may increase the PUE value.

In the past couple of years, improvements in energy efficiency of data centers have flattened out, and in some cases, even deteriorated. The biggest infrastructure efficiency gains happened over half a decade back. Further improvements will require significant investment and effort, with increasingly diminishing returns.

The industry needs performance metrics that are more accurate than PUE in measuring the efficiency of a data center. We need a clearer picture of how much power is fed to IT systems for every unit of data. One of the key limitations of PUE is that it measures the overall efficiency of the 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 the data center divided by the total input power to IT equipment. But how do you measure IT power consumption?

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 the server racks, the measurement point should be at the output of the Power Delivery Units (PDU).

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 efficiency shows how efficiently the IT equipment delivers useful output for the given electrical power input.

Site infrastructure efficiency shows 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 enable you to track efficiency over time and reveal opportunities to maximize IT output, while lowering input power by reducing losses and inefficiencies in support systems.[3]

Although there are no industry benchmarks for IT efficiency, there are some benchmarks for site infrastructure efficiency. Some of the alternative metrics to PUE include the Green Grid's Data Center Productivity (DCP) and Data Center energy Productivity (DCeP), the Uptime Institute and McKinsey's Corporate average data center efficiency (CADE), and JouleX's Performance per Watt (PPW). But for the time being, PUE remains the dominant benchmark.[4]

10 ways to improve the overall efficiency of your data center:

1. Minimize idle IT equipment

IT equipment is usually very lightly used relative to its capacity. Servers tend to be only 5 to 15 percent utilized, processors are 10 to 20 percent utilized, storage devices are 20 to 40 percent utilized, and networking equipment is 60 to 80 percent utilized.

However, even if any of these devices are idle, the equipment still consumes a significant portion of the power it would draw at maximum utilization. A typical server consumes 30 to 40 percent of maximum power even when it's producing no work at all.

Uptime Institute found that 30 percent of servers worldwide are unused. This doesn't affect the PUE of a data center but results in a loss of $30 billion in wasted electricity. To fight this, identifying underutilized pieces of equipment and powering them down can seem to be the most effective strategy.

However, while algorithms exist to turn servers on and off based on capacity needs to reduce energy consumption, turning the server off may not necessarily be the best answer. Repeated switching of servers consumes energy, along with inducing significant wear and tear of storage disks.

A better approach to deal with the 10 million "zombie" servers worldwide is distributed computing, which links computers to work as if they were a single machine. Scaling up the number of data centers working together increases their processing power, thereby reducing or eliminating the need for separate facilities for specific applications.

2. Virtualization of servers and storage

Across the industry, we can see many instances of dedicated server and storage inefficiently deployed for a single application — just to maintain physical lines of demarcation. However, with virtualization, you can aggregate servers and storage 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, are sharing the hardware with other applications. Virtualization can bring great benefits for most data centers, dramatically improving hardware utilization and enabling reduction in the number of power-consuming servers and storage devices. Virtualization can improve server use from an average of 10 to 20 percent to at least 50 to 60 percent.[5]

Virtualization is not the answer for all data centers wishing to be more efficient, particularly if they are designed to manage peak loads through the day, in which case having underutilized, idle hardware is necessary.

3. Consolidate servers, storage, and data centers

At the server level, blade servers can really help drive consolidation as they provide more processing output for the power consumed. Each blade on a single chassis shares common power supply, fans, networking, and storage. Compared to traditional rack servers, blade servers can perform the same work for 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. Use of high-speed solid-state drives should be reserved for only those applications that require instant response.

And last but not the least, if underutilized data centers could be consolidated in one location, the operator would reap great savings by sharing cooling and back-up systems to support loads, in addition to real estate savings.

Data centers' electricity demand has remained roughly level in the past five years, in part because of a shift towards 'hyperscale' facilities, which are super-efficient due to an organized, uniform computing architecture that easily scales up to hundreds 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 will fall by a quarter, if 80 percent of servers in small U.S. data centers were moved over to hyperscale facilities. Hyperscale centers already account for 20 percent of the world's data-center electricity usage and by 2020 they will draw almost half of it, the IEA says.[1]

4. Effectively manage CPU power usage

More than 50 percent of the power required to run a server is used by the 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.

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.

Such adaptive power management reduces power consumption without compromising processing capability. If a CPU is operating near maximum capacity most of the time, this feature would offer little advantage, but it can produce significant savings when CPU utilization is variable.

5. Advancements in 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 are in place to improve the efficiency of server components. 80 PLUS-certified power supplies are becoming more mainstream for modern IT equipment. The program got its name from the 80 percent minimum efficiency that a power supply needs to exhibit at 20, 50 and its full rated load.[6]

More efficient PSUs consume lower energy and generate less heat. The amount of heat produced is key, as it affects the performance of the PSU and the lifetime of all its internal parts.

PSUs of higher efficiency are expensive but lead to larger savings for the data center. If a PSU operates at 90 percent efficiency and voltage regulators operate at 85 percent efficiency, the overall energy efficiency of the server would be greater than 75 percent.

6. Opt for high-efficiency UPSs

Power drawn by a data center facility typically passes through an 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 the energy efficiency is primarily dictated by power-conversion in the UPS.

Advances in UPS technology have greatly improved their efficiency. In the 1980s, most UPS systems used silicon-controlled rectifier technology to convert battery DC power to AC power, operating at a low-switching frequency and were 75 to 80 percent efficient at best.

With the advent of insulated-gate bipolar transistor switching devices in the 1990s, switching frequency increased and power-conversion losses decreased significantly, driving up UPS efficiency to 90 percent. Today, thanks to higher-speed switches, transformers are no longer needed in UPSs, helping boost efficiency to 94 percent.

However, when evaluating a UPS, only considering the peak efficiency rating is not ideal, as they are less likely to operate under a full load. Many IT systems use dual power sources for redundancy, and a typical data center utilizes its UPSs to less than 50 percent capacity, and, in some cases, at as low as 20 to 40 percent.

Previous-generation UPSs were drastically less efficient at low loads. In addition to dramatic cost savings, today's high-efficiency UPSs extend the lifetime of batteries with better internal thermals, thereby increasing overall reliability and performance.

Moreover, most data centers use double-conversion type UPS, which converts the incoming power to DC and then back to AC. This generates a clean, consistent waveform for IT equipment, effectively isolating them from the power source. UPS systems that don't convert the incoming power—line interactive or passive standby systems—can operate at higher efficiencies.[7]

7. Distribute 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. However, most equipment is run at a lower-voltage, sacrificing efficiency for tradition.

A UPS can operate from 120-415V, and a power distribution unit (PDU) steps down that power to 208 V (IEC) or 120V (UL) to feed the servers. When that step-down transformer in the PDU is eliminated by operating IT equipment at higher voltages, the power chain can be much more efficient. By operating a UPS with 415V output power that can feed a server directly, an incremental 2 percent reduction in facility energy can be achieved.[8]

8. 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 some well-established practices.

First comes the use of hot aisle/cold aisle enclosure configuration. By arranging equipment in an alternating pattern of an aisle with a cold air intake and the next with a hot air exhaust, you can get more uniform air temperature throughout a server room. And by using blanking panels inside equipment enclosures, you can ensure that air from hot and cold aisles don't mix.

'Bypass airflow' affects as much as 60 percent of the cool-air supply in computer rooms. Sealing cable outputs minimizes this phenomenon, whereby cool air is sucked back in to cooling units, instead of circulating evenly throughout a data center.

Orienting air-conditioning units close to the enclosures and perpendicular to hot aisles maximizes cooling where it's needed most. Further optimization of cooling systems can be achieved by using efficient technologies, such as variable frequency drives for air chillers.

However, 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 2 kW to 3 kW per rack. But now, we are aspiring to create an environment that can support much higher densities, in excess of 30 kW, in which air-cooling systems falls short (air-cooling systems fall short above 25 kW).[9]

9. Operate at 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, according to current standards, this temperature can be increased by a few degrees to save energy without compromising IT service continuity and without any extra investment.

ASHRAE (the American Society of Heating, Refrigerating and Air Conditioning Engineers) has recommended rack level intake temperature ranges of 18 to 27°C and a humidity level of 8 to 60 percent since 2015. For most new devices, the recommended temperature ranges from 15 to 32°C, with a humidity tolerance of 8 to 80 percent.

Increasing rack level intake temperatures also enables the use of free cooling or free chilling systems, where outdoor air is used to blow fresh air into the room or naturally cool water is used, instead of chilled water production units.

With a recommended temperature of 25°C instead of 15°C in the server room, free cooling can be used without turning the air conditioning on for extended periods of the year. The same goes for free chilling, with recommended temperatures set at 15°C, rather than 7°C.[10]

10. **Participate with the Smart Grid**

Smart grids, the next generation power grids, enables two-way energy and information flows to create an automated and distributed power delivery network. Whereas traditional electrical grids are mainly focused on just carrying power from several central power generators to the customers, smart grids employ real-time monitoring devices to minimize disruptions.

They can automatically detect problems, immediately respond to errors in power lines, and accurately isolate error-prone links from the main power network. Communication and metering technologies of smart grids can inform data centers via smart devices when power demand is high in their regions, a massive advantage in an era of dynamic pricing.

Smart grids can also 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.

Data center operators can not only draw clean power from the grid, but they can also install renewable power generators at the 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.

Conclusion

Data centers use about two percent of the world's electricity. Burning over 416-terawatt hours of electricity each year, today’s data centers support an ever-growing information and communications technology sector, with a carbon footprint at par with the aviation industry. If the computationally intensive cryptocurrency usage and advancement of 5G technology gains wider popularity as expected, energy consumption will grow exponentially.

To manage this expected surge, data centers need to become more energy-efficient. Tackling IT efficiency holds most promise as optimal usage of resources, through both better hardware and software, not only brings down power consumption but also increases the actual computational output of the facility, a factor often overlooked in the race for better PUE figures.

Another promising area is tweaking of cooling infrastructure. As the industry continues its march towards higher rack density, the future might increasingly involve liquid cooling technologies. In the meantime, some simple equipment reorganization and better facility management to take advantage of greater temperature tolerance of modern gear can greatly improve data center cooling, without new investment in hardware.

More efficient UPSs and PSUs, combined with equipment capable of operating at higher voltages, will help datacenters lower their energy consumption. They will also be able to manage power more effectively, with the widespread adoption of smart grid technologies and on-site renewable sources to generate electricity. We are moving towards a future where data centers are not just consumers but also contributors to power grids.

References

1. https://www.iea.org/digital/
2. https://uptimeinstitute.com/uptime_assets/35f30b159b6ecd6ce38aa2bc71d5d34a4939438fe51ad5b7cbc323f4d6aaaf26-2019-data-center-industry-survey.pdf
3. https://www.missioncriticalmagazine.com/ext/resources/MC/Home/Files/PDFs/WP_LinkedIN%20DataCenterEnergy.pdf
4. https://www.osti.gov/servlets/purl/1372902
5. https://www.datacenterknowledge.com/archives/2008/09/01/virtualization-to-increase-business-efficiency-and-decrease-costshttps://www.nisg.org/files/documents/E03010005.pdf
6. https://www.tomshardware.com/news/what-80-plus-levels-mean,36721.html
7. https://ashraephilly.org/images/downloads/CTTC_Articles/1503_cttc.pdf
8. https://www.energy.gov/sites/prod/files/2013/10/f3/eedatacenterbestpractices.pdf
9. https://tc0909.ashraetcs.org/documents/ASHRAE_TC0909_Power_White_Paper_22_June_2016_REVISED.pdf
10. https://www.apl-datacenter.com/en/increasing-server-room-temperature-to-improve-data-center-energy-performance/

ABB
3055 Orchard Dr
San Jose, CA
United States

abb.com/datacenters

9AKK107492A6219