White paper

The two sides of highly efficient data centers

The typical data center has a voracious energy appetite, using thirty times more energy per square foot than most office buildings. Little wonder one of the major constraints on increasing data center capacity is insufficient power.

The other major constraint is physical space. The need for more computational power is often resolved by adding servers. Inevitably, though, rack space is maxed out and data center managers must find new ways to expand capacity.

Operators can tackle both constraints by focusing their attention on enhancing the efficiency of their data centers. Through changes to how they equip and operate their centers, it's possible to greatly enhance efficiency and squeeze more transactions at a lower cost from every square foot of data center real estate.

Measuring Efficiency

A mantra of continuous improvement experts is that you can't improve what you can't measure. Crude measures like total-power-in/month were usually used to gauge data center energy efficiency. This ratio provides a big-picture view, but not the granular, real-time data needed to implement efficiency improvement. As in sports, knowing the score at the end of the game doesn't help the team improve performance. More precise and timely performance data is required.

"Power Usage Effectiveness (PUE) is one measure of how efficiently a data center uses power," says Clemens Pfeiffer, Chief Technology Officer at Power Assure. "It is the ratio of the power actually delivered to the racks divided by total-power-in. The more power used to meet facility requirements, primarily cooling, rather than power servers, the lower the efficiency."

While potentially providing a clearer view of efficiency, PUE figures are frequently inaccurate because facility usage data often isn't based on measurement. Rather, the calculations rely on nameplate data on servers and other devices. Actual power used by a given device may be as much as 65% less than what's shown on the nameplate because manufacturers tend be very conservative. They err on the side of indicating higher usage to ensure that customers build in a margin of safety when powering the equipment.[1]

More accurate and useful PUE data requires having device-level instrumentation to report real-time, actual energy consumption. Realizing that data center operators are increasingly concerned about energy use, manufacturers are designing smarter equipment with the instrumentation required to report usage data. Older, existing equipment can also be retrofitted with metering instruments.

While PUEs are a step in the right direction, there are still better efficiency metrics available.

"Performance per watt/second is a more precise and useful metric," Pfeiffer says. "It is a measure of the energy efficiency of a particular computer architecture or computer hardware - the rate of computation delivered for every watt/second of power. This level of detailed measurement is mostly lacking for servers today."

1. http://www.powerassure.com/images/pdfdocs/collateral/netapp_story_in_searchdata-center.pdf

Underwriters Laboratories recently announced their implementation of this metric, UL 2640, saying that it "will help data centers more accurately determine and optimize the utilization of server capability." Based on Power Assure's PAR[4] methodology, UL 2640 provides a way to compare efficiency between servers regardless of age or capacity. Operators can use this information to better assess server utilization and power consumption, and to optimize center efficiency.[2]

Once a data center has some means and metrics to measure efficiency gains, it's possible to begin implementing improvement. Potential efficiency improvements lie in three broad areas: Asset utilization, operational efficiency and facility energy efficiency. Most people focus just on facility energy efficiency, but that's typically a one time effort to maximize power availability for IT equipment. It's much more productive long term to devote your attention to using assets and operating the facility more efficiently. With success in those areas, overall energy efficiency will naturally follow.

Boosting Asset-Based Efficiency

Data center mangers need to be prepared fc
possible volume of transactions.

"If you are a news website, you need to be ready to handle the transactions when a known musician dies and everyone hits your site to learn more," Pfeiffer says. "Data center planners provision IT equipment based on those expected demands. Then they take that capacity and replicate it in a second data center for redundancy. Of course, at 3:00 a.m. on a day when there's no big news, the demand is far, far less. Still, the capacity is there and usually online."

It's like owning a landscaping company that usually delivers dirt by the pickup load. The company receives the occasional large order that requires a dump truck, so that's the vehicle the company owns and operates full time. Clearly, that's a waste of capacity and resources, but it's necessary to have that capacity to deliver the large orders.

"When you take that kind of approach, it's little wonder that the average utilization on the data center is around 10%," Pfeiffer explains. "They need to have that capacity available; the question is how to most-effectively manage that capacity. We're not doing that very well today."

The amount of available energy is the primary constraint on increasing capacity in data center. (This topic is addressed later in this white paper.)

"The second, and sometimes just as important, constraint to increased capacity is the physical space for racks and servers," Pfeiffer says. "I've visited data centers where each rack has something like 10,000 watts of power. Because of the energy needs of their servers, they can only power two servers per rack, so each rack is only half filled. They may have overbuilt their facility in terms of space, under provisioned the racks in terms of power or they picked the wrong equipment. Regardless of the reason, the result is a lot of expensive space going unused."

Another way rack real estate is underutilized is grouping racks or servers by application or customer. This makes sense from on an operational standpoint because it's easier to quickly find a specific server. However, the more efficient approach would be to organize servers based on space and power available. This lets you squeeze additional equipment into the data center. It's more complicated operationally to have servers for a given customer or application spread throughout the data center, but a spreadsheet or chart can guide technicians to a specific server.

"You want to maximize the utilization of the power and space so that you don't end up with half empty racks or circuits that are only 50 or 60% utilized," Pfeiffer says.

Direct Efficiency Improvement

You can't fight physics. The unalterable fact is that any time you e the direction or form of energy, some of it leaks away and is lost. Change the direction of RPM of a shaft in a machine by running it through a gearbox and you've lost some of the original power.

In a data center, electric energy goes through a number of conversions between the street and the server. The incoming 480 VAC is rectified to DC to charge the batteries in the UPS.[3] The output from the UPS is inverted back to AC and distributed throughout the center. In many systems, there is also an input transformer used to electrically isolate UPS. Power distribution units then step the voltage down before it's delivered to the racks. There, it's converted inside each server back to DC.

Data center managers are increasingly coming to the realization that converting AC current in each server is tremendously inefficient.

"Since the servers all use DC power, it makes sense to convert AC to DC once, and have a DC bus that directly supplies the energy to the servers," says Gary Rackliffe, the ABB vice president of smart grids.

"The concept that we're moving toward is, instead of having the individual pieces of equipment, like servers and processors convert AC to DC, doing this at a higher level and distributing DC," says Mietek Glinkowski, Director of Technology at ABB.

Converting to DC power not only delivers energy with reduced waste, it also improves asset efficiency because there is less equipment required. The footprint of electrical equipment in a

2. http://www.ul.com/global/eng/pages/corporate/aboutul/publications/newsletters/high-tech/vol2issue3/4par/
3. Rackliffe: "Do data centers actually feed power under normal conditions through a UPS with AC to DC and DC to AC conversions?"

DC data center can be reduced by 30% or more compared to AC.[4] This allows planners to build smaller centers or increase the transactions per square foot.

From a capital standpoint, DC power also provides a big benefit. DC-powered data centers are 29% less expensive than AC centers in terms of average equipment and installation costs.[5]

As an added operational efficiency, DC data centers create 30% less heat than comparable AC data centers because DC centers require fewer power conversions.[6] That decreases the load – and therefore the operating costs – of the cooling system.

The efficiencies are large and confirmed, hence the growing interest in rewiring for DC.

Naturally, center planners are hesitant to stray from a power standard that's been in place for a century. They were also held back by the unavailability of DC-equipped servers and related hardware. Even the plugs have to be changed at the rack. Manufacturers, seeing the growing demand for DC solutions, are ramping up production of DC-powered devices.

Virtualization

One of the most common strategies for maximizing the utilization, and therefore the efficiency, of servers is virtualization. In this approach, specific servers aren't dedicated to a particular application. Instead, the system searches for available capacity on any one of a number of servers in a pool and doles out the transactions accordingly.

"Today, you typically have one application running on that server," Pfeiffer says. "With virtualization, the physical location of the server becomes independent of the business unit or application. You can take a logical view of computing resources rather than a physical view. That makes it easier to arrange servers in a way that makes the most sense in terms of power and space."

There are many approaches to virtualization. Some installations virtualize existing servers, adding the ability to move transactions or applications. This enhances reliability because any server can step in when another fails. However, unless under-used servers are powered down, there's no efficiency gain.

"To improve efficiency, you usually rely on fewer, more-powerful servers," Pfeiffer says. "You can go from 100 servers to 10 because of the performance improvement of the hardware. Virtualization allows you to run 100 independent applications on these 10 physical servers as if they were 100 individual servers. That's where the efficiency increase comes from."

Consolidation of assets through virtualization creates a seamless server pool, with the ability to more readily shift applications and transactions. This can improve efficiency by shrinking the scale of future hardware expenditures and freeing valuable rack space. Little wonder that virtualization is high on the hit list of efficiency improvement strategies for many people.

"Industry analysts report that between 60 percent and 80 percent of IT departments are pursuing server-consolidation projects. By reducing the numbers and types of servers … companies are looking at significant cost savings."[7]

Operational efficiency

The changing energy landscape is encouraging utilities to create new and inventive ways to better manage power consumption. Like data centers, electric utilities also face increased overall demand and deal with the huge daily swings in demand level.

"During peak times, utilities ramp up generation to meet increased demand," says Rackliffe. "During critical peak demand periods, they will bring additional generation facilities online, and sharply ratchet up the price for electricity during those periods. The last generation assets to be dispatched are typically the older, less-efficient and higher-polluting. Reducing critical peak demands can save both operational costs and capacity costs."

In years past, electric utilities promoted increased consumption to enhance revenues. Today, utilities have incentives to avoid additional capital expenditures, and penalties from regulators like the EPA. The utilities actively encourage customers to reduce the number of kilowatts consumed and to balance their use of the power that is consumed. One of the most powerful ways to balance loads is variable pricing.

The End of Fixed-Price Power

"Many utilities are now charging in unique ways," Pfeiffer says. "Power costs are no longer at a flat rate of, say $0.07 per kilowatt/hour. That cost may triple during peak periods."

Utility power contracts can include a peak demand limit, which is a powerful motivator for the data center to avoid usage spikes.

"In these agreements, part of the bill for the entire contract period is based on the data center's peak demand for any one time interval throughout the period," Rackliffe says. "A single time period with increased usage can make a huge difference in total annual energy costs."

Depending on the configuration and flexibility of the facility and network, data center operators may be able to capitalize on several approaches to avoid paying top-dollar for their power.

4. ABB PPT
5. ABB PPT
6. ABB PPT
7. Virtualization Definition and Solutions, CIO Magazine, http://www.cio.com/article/40701/Virtualization_Definition_and_Solutions

"Many centers take advantage of the geographic redundancy of their network," Pfeiffer says. "When the price of energy peaks in their downtown Manhattan facility, they will shift capacity to centers farther west where the cost is relatively lower."

Geographically relocating the loads only creates added efficiency, however, if the center shifting the load has the ability to throttle back on consumption. Despite the enormous computational power of the average server, in one very important way it is not as smart as a chipmunk. At rest, the chipmunk burns relatively little energy. Its metabolism ramps up when needed to search for food and peaks when evading a cat. Most servers, on the other hand, burn energy at a constant rate. Whether idling or working full bore, the energy consumed remains about the same.

The ability to automatically power down unused servers is a feature increasingly being incorporated in Data Center Infrastructure Management (DCIM) software. While DCIMs were mainly an information source in the past, they are now becoming true management tools providing the guidance and automation to intelligently shift loads and reduce power use.

The Getting-Smarter Grid

The smart grid increasingly lives up to its name and offering new efficiency potential.

"As it relates to the data center, one of the key concepts of the smart grid is the tighter communication with the utility company," Pfeiffer explains. "Both parties realize that it's in their best interests to cooperate, reducing peak demand on the utility while ensuring reliable, lower cost power to the data center. That is the kind of thing the smart grid lets data center managers do… leverage for price and risk mitigation."

The utility or demand response integrators may negotiate demand response agreements with the data center operators. In these agreements, incentives are offered in return for the promise that the data center will not exceed certain usage levels or will take loads offline when requested in order to reduce demand. The agreements typically include defined durations and limits on number of requests by the utility or integrator.

"In this scenario, the data center can make dual use of its backup generation capability," Rackliffe says. "Backup generators become a cost-management tool. These assets are no longer used only to compensate for lost power from the grid. They are also used to proactively to reduce demand from the grid while maintaining the power needed. This can reduce your spend on energy by helping set a lower, annual rate."
Diesel and other methods of backup generation typically aren't highly efficient, so center managers need to consider the cost/benefit of those higher costs compared to paying higher per/kW utility costs.

Unified View Creates Greater Efficiency

Senior leaders in many organizations talk about the concept of "aligning incentives with business goals." What that means is making sure that employees are encouraged to do what's best for the organization as well as what's best for them.

In the data center, there are two primary factions with sometimes-competing incentives. The facilities staff is responsible for the physical building, power plant and HVAC. They will be rewarded for reliable operation with minimal downtime, as well as for minimizing costs, mainly energy.

The IT group, on the other hand, is the keeper of the data, interested almost exclusively in ensuring that its flow in and out of the racks is uninterrupted. How much energy it takes to do that is of secondary concern.

"Whatever energy efficiency targets you put in place, they will be secondary concern to the IT group," Pfeiffer says." No IT person ever got fired for using too much electricity, but if there are a number of service interruptions, it's likely that heads will roll."[8]

Because of these somewhat misaligned incentives, these data center staff members sometimes work at cross-purposes.

The Unifying Power of DCIM

By integrating the IT and facility management disciplines in a centralized monitoring and management application, a DCIM can help center staff work cooperatively toward a tightly aligned objective of overall efficiency. Supported with the proper sensors and network, DCIM enables a common, real-time monitoring and management platform.

"There may be 10 or more measurement or management systems in place," says Pfeiffer. "You can't look at only one of them and try to maximize its efficiency. You have to simultaneously look at power, cooling, server utilization and more. There are interdependencies between all of these values. If you only look at one, that's not sufficient or intelligent. You may do things to optimize the efficiency of that system that will reduce efficiency elsewhere.[9]

The "single information window" provided by a DCIM provides actionable, real-time insight into energy usage, utilization and other key performance indicators (KPIs) for data centers, as well as the means to analyze and automate energy efficiency while maintaining application service levels.

Having the right instrumentation is a must-have to gain a clear view of current operations and measure incremental improvement.

8. Clemens
9. Clemens

“Most data centers have some kind of instrumentation,” says Pfeiffer. “Some of it may be focused on the facility or operational performance, while other sensors measure IT-related performance. The DCIM can aggregate that data to create comprehensive, dynamic reporting.”

The commercial-grade instrumentation being used by many data centers is not built for the high level of accuracy needed to guide efficiency improvement activities. A single cooling system temperature sensor that is off by one degree could cost a large data center upwards of $100,000 each year. Industrial-grade instrumentation, with accuracy as much as ten times more accurate than commercial equipment, gives operators a clearer view of their processes and allows them to discriminate subtle efficiency enhancements.[10]

Configured properly, DCIM can drastically improve server utilization between racks and between facilities, increasing server utilization more than 70%. This has been proven to reduce electrical usage by as much as 50% while delivering near 100% availability.[11]

The information provided by DCIM empowers the operator to shift loads to keep availability and quality of service high. It also helps data centers actively monitor energy usage and respond quickly when needed to reduce loads and avoid establishing a new peak.

The tremendous potential of DCIM as an efficiency improvement tools explains the widespread adoption of this technology. Gartner predicts sales of DCIM tools will grow 60% by 2014.[12]

Conclusion

The average data center manager has a different concept of energy efficiency than most other consumers, whether they be residential, commercial or industrial. For most energy users, efficiency means reducing how much they spend on power by wasting less of it.

“At the end of the day, most data centers don‘t really want to save energy that way,” Pfeiffer concludes. “They want to waste less power so it can be used differently… productively. Most data centers operate at, or are approaching, their capacity, with the constraining factor being the amount of power available. Reducing the wasted energy allows operators to increase the capacity of their existing centers and get the fullest return on their investment while meeting the demand from their customers.”

ABB Data Centers
125 East County Line Road
Warminster, Pennsylvania, U.S.A.
Phone: +1 800 HELP 365

www.abb.com/datacenters

9AKK105408A8197

10. ABB PPT
11. ABB PPT
12. http://searchcio.techtarget.com/news/2240035504/NASAs-IT-redo-takes-off-guided-by-tools-for-data-center-efficiency