Virtualization and Cloud Computing: Optimized Power, Cooling, and Management Maximizes Benefits

White Paper 118

Revision 4

by Suzanne Niles
Patrick Donovan

> ## Executive summary

IT virtualization, the engine behind cloud computing, can have significant consequences on the data center physical infrastructure (DCPI). Higher power densities that often result can challenge the cooling capabilities of an existing system. Reduced overall energy consumption that typically results from physical server consolidation may actually worsen the data center's power usage effectiveness (PUE). Dynamic loads that vary in time and location may heighten the risk of downtime if rack-level power and cooling health are not understood and considered. Finally, the fault-tolerant nature of a highly virtualized environment could raise questions about the level of redundancy required in the physical infrastructure. These particular effects of virtualization are discussed and possible solutions or methods for dealing with them are offered.

Contents

Click on a section to jump to it

Schneider Electric

Introduction

Without question, IT virtualization - the abstraction of physical network, server, and storage resources - has greatly increased the ability to utilize and scale compute power. Indeed, virtualization has become the very technology engine behind cloud computing itself. While the benefits of this technology and service delivery model are well known, understood, and increasingly being taken advantage of, their effects on the data center physical infrastructure (DCPI) are less understood. The purpose of this paper is to describe these effects while offering possible solutions or methods for dealing with them.

These effects are largely not new and successful strategies for dealing with them exist today. There are four effects or attributes of IT virtualization that will be covered in this paper.

1. **The rise of high density** – Higher power density is likely to result from virtualization, at least in some racks. Areas of high density can pose cooling challenges that, if left unaddressed, could threaten the reliability of the overall data center.

2. **Reduced IT load can affect PUE** – After virtualization, the data center's power usage effectiveness (PUE) is likely to worsen. This is despite the fact that the initial physical server consolidation results in lower overall energy use. If the power and cooling infrastructure is not right-sized to the new lower overall load, physical infrastructure efficiency measured as PUE will degrade.

3. **Dynamic IT loads** – Virtualized IT loads, particularly in a highly virtualized, cloud data center, can vary in both time AND location. In order to ensure availability in such a system, it's critical that rack-level power and cooling health be considered before changes are made.

4. **Lower redundancy requirements are possible** – A highly virtualized data center designed and operated with a high level of IT fault-tolerance may reduce the necessity for redundancy in the physical infrastructure. This effect could have a significantly positive impact on data center planning and capital costs.

This paper approaches these effects in the context of a highly virtualized environment typical of a cloud-based data center with dynamic demand requirements. The list of white papers at the end provides additional general and detailed information about these topics in the overall data center context, virtualized or not.

The rise of high density

While virtualization may reduce overall power consumption in the room, virtualized servers tend to be installed and grouped in ways that create localized high-density areas that can lead to "hot spots". This cooling challenge may come as a surprise given the dramatic decrease in power consumption that is possible due to high, realistically achievable physical server consolidation ratios of 10:1, 20:1 or even much higher. As a physical host is loaded up with more and more virtual machines, its CPU utilization will increase. Although far from being a linear relationship, the power draw of that physical host increases as the utilization increases. A typical non-virtualized server's CPU utilization is around 5%-10%. A virtualized server, however, could be 50% or higher. The difference in power draw between 5% and 50% CPU utilization would be about 20% depending on the specific machine in question. Additionally, virtualized machines will often require increased processor and memory resources which can further raise power consumption above what a non-virtualized machine would draw. Grouping or clustering these bulked up, virtualized servers can result in significantly higher power densities that could then cause cooling problems. Not only are densities increasing, but virtualization also allows workloads to be dynamically moved, started, and stopped – the result can be physical loads that change both over time AND in their physical location. This effect of dynamic loads is discussed later in this paper.

Methods for cooling high density racks to prevent "hot spots"

Higher rack power densities should encourage data center operators to examine their existing cooling infrastructure to see if it can still sufficiently cool the load. Several approaches for cooling high density racks exist. Two of the principal ones are briefly explained here. Perhaps the most common method is to simply "spread out" the high density equipment throughout the data center floor rather than grouping them together. By spreading out the loads in this way, no single rack will exceed the design power density and consequently cooling performance is more predictable. The principle benefit of this strategy is that no new power or cooling infrastructure is required. However, this approach has several serious disadvantages including increased floor space consumption, higher cabling costs, possible reduced electrical efficiency related to uncontained air paths and the perception that half-filled racks are wasteful. That being said, this simple approach may be a viable option particularly…

- When the resulting average data center power density (kW/rack or watts/sq foot of white space) is about the same or less than it was before virtualization. (Assumes there was sufficient cooling capacity before.)
- When managers have complete control over where physical servers are placed.
- When "U" space is available in existing racks to allow the spreading to happen.

A more efficient approach may be to isolate higher density equipment in a separate location from lower density equipment. This high density pod would involve consolidating all high density systems down to a single rack or row(s) of racks. Dedicated cooling air distribution (e.g., rack and row-based air conditioners), and/or air containment (e.g., hot or cold aisle containment), could then be brought to these isolated high density pods to ensure they received the predictable cooling needed at any given time. The advantages include better space utilization, high efficiency, and that it enables maximum density per rack. Additionally, for organizations that require high density equipment to remain co-located together, this approach is obviously preferred. **Figure 1** illustrates the idea of a high density pod.
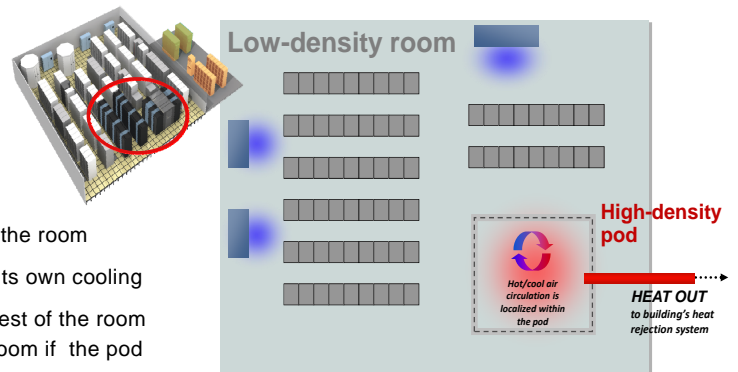
Figure 1

The high-density "pod" is an option for dealing with high density equipment in an existing, virtualized data center

- A high-density "island" in the room
- A "mini data center" with its own cooling
- Thermally neutral to the rest of the room (may actually assist the room if the pod has excess capacity)
- Hot/cool air circulation is localized within the zone by short air paths and/or physical containment

Important basic characteristics of a high density pod include:

- Short air paths (when air streams are uncontained) to minimize possibility of mixing supply and return air, as well as to minimize energy used by variable speed fans.
- Variable Frequency Drive (VFD) fans to account for dynamic loads that vary in both time and location.

- Rack and/or row orientation that facilitates the separation of cold supply air from hot return air.

Isolating high density equipment into a separate pod with dedicated cooling makes management simpler, cooling performance more predictable and energy is used more efficiently. At a minimum, the high density pod is thermally neutral to the rest of the room. But, particularly when air streams are well contained, the pod may likely add cooling capacity to the rest of the data center.

For more detailed information regarding the use of separate, isolated high density pods, see White Paper 134, *Deploying High Density Pods in a Low Density Data Center*. For more general information about all the various methods for cooling high density equipment, see White Paper 46, *Cooling Strategies for High Density Racks and Blade Servers*.

Reduced IT load can affect PUE

A widely touted benefit of virtualization has been reduced energy costs as a result of physical server consolidation. And, indeed, these savings are often not trivial. Consider a 1MW data center with 1000 physical servers that draw 250W each at a cost of $0.11 per kilowatt/hr. The energy cost per year to operate just the servers would be about $240,000 (250W/1000 x 0.11 x 24hr x 365 days x 1000 servers). Virtualize these servers at a conservative consolidation ratio of 10:1 with each remaining physical server operating at a CPU utilization of 60% (instead of the typical 5-10%) and you now get a total energy cost of about $60,000 (600W/1000 x 0.11 x 24hr x 365 days x 100 servers)[1]. That represents an energy savings of 76% for the servers. So it is no wonder that virtualization is widely viewed and promoted as a "green" technology for the data center. Compute capacity remains the same or is even increased while energy use drops sharply.

It may come as a surprise, then, that in such a "green" scenario the most commonly used metric for data center efficiency, PUE, could worsen after this server consolidation took place. Perhaps some might see this as a short-coming of the metric. That is, an efficiency metric must be deficient if it is not intended to reflect the obvious environmental benefit of significantly lower energy use. But the reader must remember that PUE is a metric designed to only measure efficiency of the data center *physical infrastructure* (i.e. power and cooling), and not the IT compute power efficiency. PUE should never be used or thought of as a direct indicator for how "green" a particular data center is. The purpose is to show how efficient the power and cooling systems are for a given IT load. **Figure 2** illustrates the effect of virtualization on data center efficiency.

Figure 2

Typical effect of virtualization on data center efficiency showing context of PUE within overall data center efficiency


---
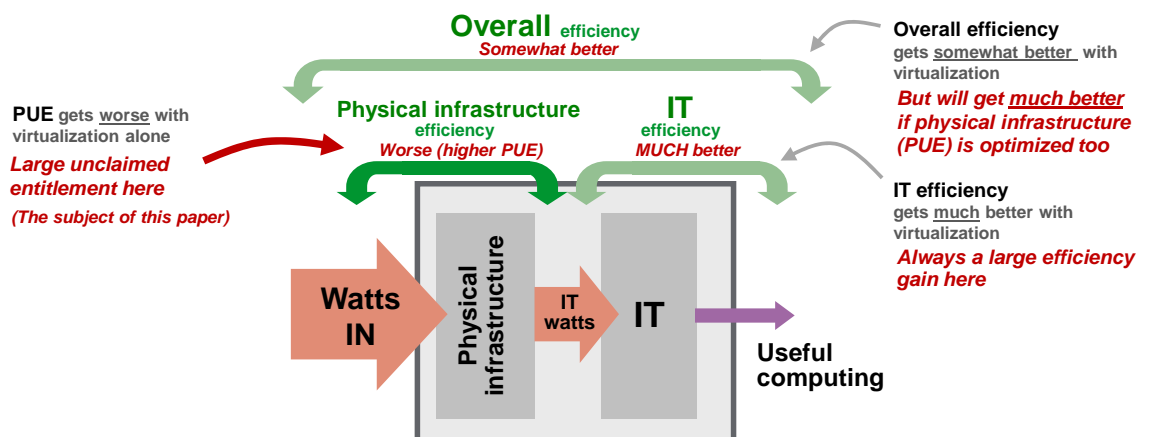
[1] The same physical host at 60% CPU utilization would bring the power consumption to about 300W. The additional 300W to bring the total to 600W is assumed due to the increased processor and memory resources required to properly handle the virtualized loads. Note that storage energy from NAS or SAN may increase but overall IT energy savings is still significant.

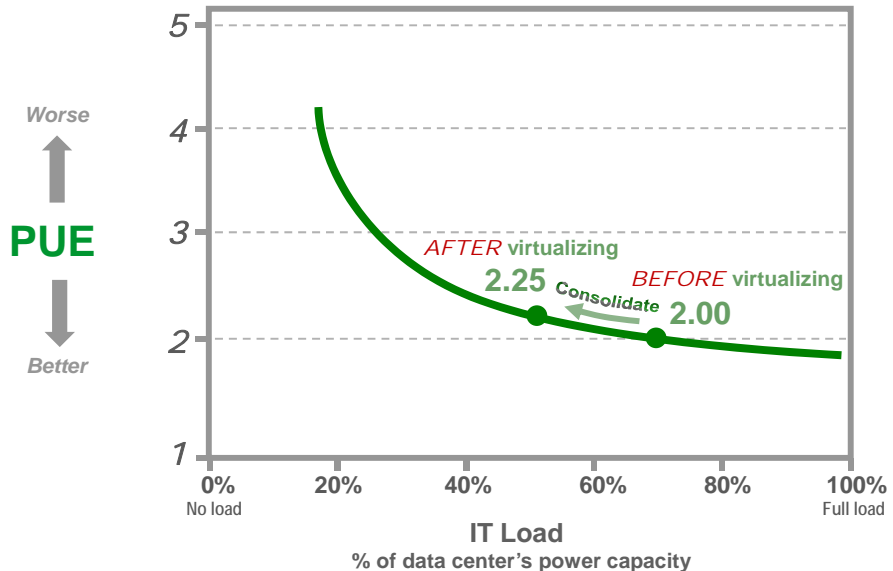## Virtualization's track on the infrastructure efficiency curve

If the power and cooling infrastructure is left alone as it was before virtualization was implemented, then PUE will worsen after the physical consolidation of servers and storage has taken place. Inherent in unused power and cooling capacity is what is known as "fixed losses". This is basically power that is consumed by the power and cooling systems regardless of what the IT load is. The more power and cooling capacity that exists, the more fixed losses will exist. As the IT load shrinks (e.g., from consolidation) these fixed losses become a higher proportion of the total data center energy use. This means PUE will worsen. This also means that PUE is always better at higher IT loads and worse at lower loads. **Figure 3** shows a typical PUE curve illustrating the relationship between efficiency and the IT load.

Figure 3

Typical data center infrastructure efficiency curve
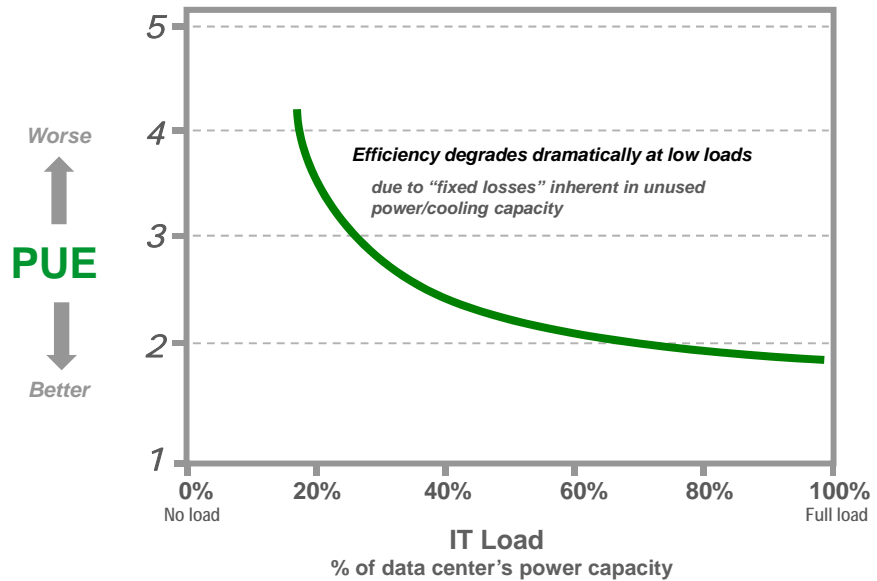
Each data center will have a higher or lower PUE curve – depending upon the efficiency of its individual devices and the efficiency of its system configuration – but the curve always has this same general shape. **Figure 4** shows the effect of consolidation on PUE.

Figure 4

Consolidation reduces load, moving PUE up the curve

NOTE: These PUE numbers are based on a traditional data center with raised floor cooling, perimeter-based constant speed cooling units, and non-modular UPS.

To improve post-virtualization PUE, the data center's infrastructure efficiency curve must be improved (lowered) by optimizing power and cooling systems to reduce the waste of oversizing and better align capacity with the new, lower load (**Figure 5**). In addition to improving efficiency, optimized power and cooling will directly impact the electric bill by reducing the power consumed by unused power and cooling capacity. The case study used to quantify these PUEs is located the **Appendix** of this paper.

Figure 5

Optimized power and cooling improves the post-virtualization efficiency curve
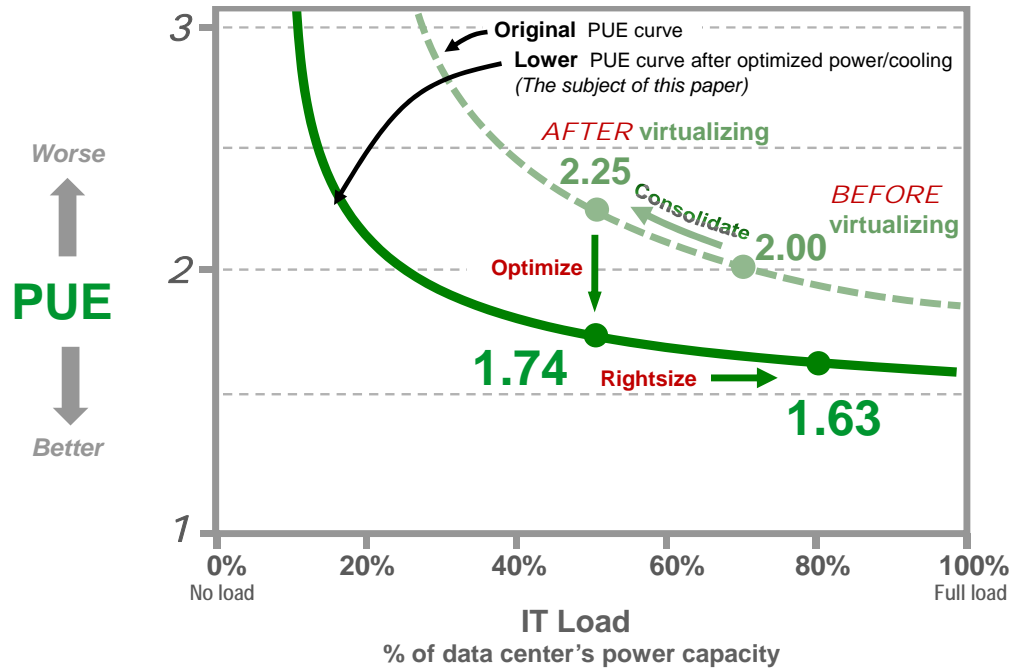
Link to resource
White Paper 113

Efficiency Modeling for Data Centers

For more information about efficiency as a function of load, see White Paper 113, *Efficiency Modeling for Data Centers*.

To improve PUE, reduce fixed losses

To realize the full energy-saving benefits of virtualization, an optimized power and cooling infrastructure should incorporate design elements such as the following to minimize fixed losses and maximize the electrical efficiency of the virtualization project overall:

- Power and cooling capacity scaled down to match the load (e.g. turn off some cooling units or remove UPS modules from scalable UPS)
- VFD fans and pumps that slow down when demand goes down
- Equipment with better device efficiency, to consume less power in doing the job
- Cooling architecture with contained or shorter air paths (e.g. contain the hot or cold aisle or move from perimeter room-based to row-based)
- Capacity management system, to balance capacity with demand and identify stranded capacity
- Blanking panels to reduce in-rack air mixing of exhaust air with cold supply air

Although likely to have the biggest impact, scaling down power and cooling capacity may be the most difficult to implement for an existing data center. Reducing the capacity of these systems simply may not be feasible in certain situations. The system may not be divisible if the design is not modular. The infrastructure systems in question could be new and recently paid for, making replacement or significant modification impractical. The dependence of

other systems or operations outside of the white space on a given infrastructure element (e.g. packaged chiller plant) may preclude the time required and ability to reduce its capacity. So the costs of rightsizing the power and cooling infrastructure may very well outweigh the benefits of closely aligning supply with demand. For an existing data center, perhaps more feasible options might include doing such things as:

- Installing blanking panels to reduce air mixing
- Orienting racks into separate hot and cold aisles
- Install air containment technologies
- Adjusting fan speeds or turning off cooling units
- Removing unneeded UPS power modules for scalable UPSs

For a new data center that is currently in design, however, rightsizing the entire power and cooling plant to the load makes more sense. Doing so at this phase will mean lower initial upfront capital costs and much better energy efficiency once the data center is operational. As the example in the appendix illustrates, an existing data center can still see significant energy savings and PUE improvement even if they only focus on making their existing power and cooling systems operate more efficiently by implementing the options listed above.

Standardization and modularity make rightsizing easier

However, in spite of the feasibility issues just described for existing data centers, oversized capacity is inefficient and wasteful. And as the sidebar to the left shows, there are additional negative effects from having an under-loaded power and cooling system. Fortunately, today's power and cooling solutions make it easier even for an existing data center to keep capacity more closely aligned with actual demand. This is, in large part, because they are scalable for capacity. And this scalability is derived from both standardization and modularity. Adding capacity to traditional, non-modular systems required onsite engineering to custom design and assemble various components together from multiple manufacturers. Later modifying or reducing capacity for those traditional systems would be costly, highly disruptive and time-consuming. By designing power and cooling systems so that they are standardized, factory assembled/tested and modular in nature, it becomes much easier and less risky to add or remove capacity. For example, UPS power capacity can be quickly and safely reduced by simply removing power modules from the chassis. This can be done without interrupting power to the load and specialized service personnel are not required. Removing these power modules reduces fixed losses, thereby improving PUE. These modules can be set aside and later reused as the load increases over time. Cooling systems today are often scalable as well. Variable speed fans, for example, can increase or decrease speed as needed based on the heat load. As the overall IT load decreases after initial virtualization, fans can run at a much lower speed thereby reducing proportional losses (power consumed by the system that is proportional to its load) which also helps improve PUE. Scalable physical infrastructure closely aligned with real-time management (to be discussed later) can help provide the right amount of power and cooling where it's needed, when it's needed. This makes rightsizing a more realistic option once a data center has been virtualized.

So how much additional energy savings can be had by rightsizing the infrastructure?

TradeOff Tool™ for calculation of virtualization savings

Figure 6 shows the Virtualization Energy Cost Calculator TradeOff Tool™. This interactive tool illustrates IT, physical infrastructure, and energy savings resulting from the virtualization of servers in a data center. The tool allows the user to input data regarding data center capacity, load, number of servers, energy cost, and other data center elements.

> Effects of extreme under-loading

Unless power and cooling are down-sized to bring loading back within normal operating limits, these effects could result in expenses that negate some of the energy savings or, in some cases, pose a risk to availability.

Cooling (too-low thermal load)
- Safety shutdown because of high head pressure on compressors
- Short-cycling of compressors from frequent shutdown, which shortens compressor life
- Possible voiding of warranty from consistent operation below low-load limit
- Cost of hot-gas bypass on compressors to simulate "normal" load to prevent short-cycling

Generator (too-low electrical load or too many generators)
- Unburned fuel in the system ("wet stacking"), which may result in pollution fines or risk of fire
- Cost of unneeded jacket water heaters to keep engines warm
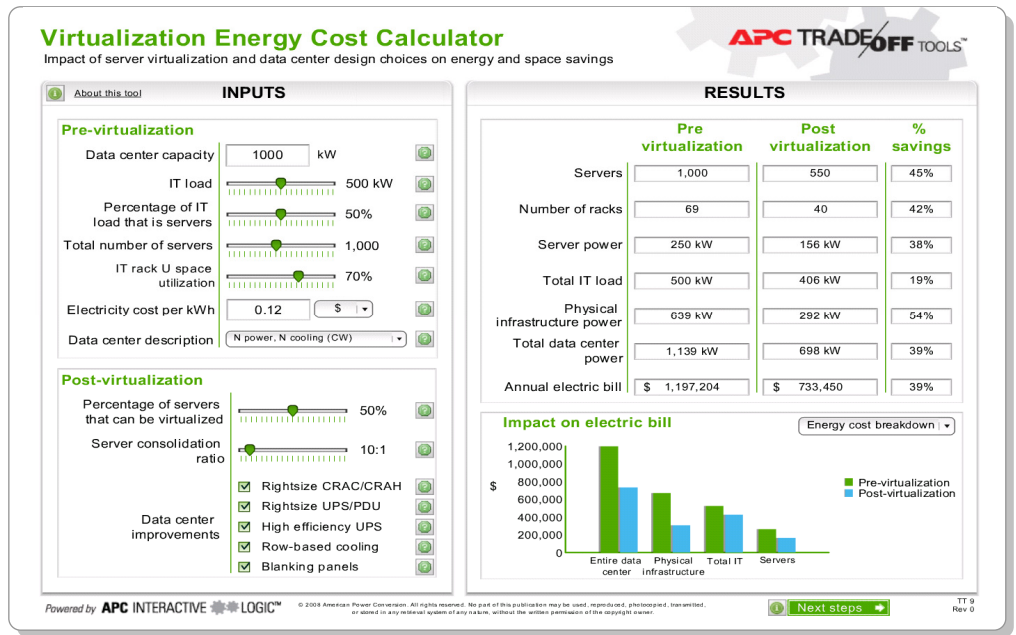- Cost of storage, testing, and maintenance of excess fuel

---

Figure 6

TradeOff Tool™ for calculating virtualization savings

Click on the screen image to link to a live version of this interactive tool

> Effect of reduced power consumption on energy and service contracts

An abrupt reduction in power consumption may have unintended consequences with regard to utility and service contracts. Such contracts will need to be reviewed and renegotiated where necessary, in order not to forfeit data center savings to the utility provider, building owner, or service provider.

- Utility contract – Agreements with utility providers may include a clause that penalizes the customer if overall electrical consumption drops below a preset monthly consumption amount.
- Energy clause in real estate agreement – Some real estate agreements include the cost of electricity as a flat rate, typically billed on a cost-per-square-foot basis. This agreement may need to be renegotiated, otherwise the savings from virtualization will accrue to the building owner.
- Equipment service contracts – Service contracts should be reviewed to remove unused power and cooling equipment, to avoid paying for service on equipment that has been taken out of service through down-sizing.

Dynamic IT loads

The sudden - and increasingly automated - creation and movement of virtual machines requires careful management and policies that contemplate physical infrastructure status and capacity down to an individual rack level. Failure to do so could undermine the software fault-tolerance that virtualization imbues to cloud computing. Fortunately, tools exist today to greatly simplify and assist in doing this.

The electrical load on the physical hosts can vary in both time and place as virtual loads are created or moved from one location to another. As the processor computes, changes power state or as hard drives spin up and down, the electrical load on any machine – virtualized or not - will vary. This variation can be amplified when power management policies are implemented which actively power machines down and up throughout the day as compute needs change over time. The policy of power capping, however, can reduce this variation. This is where machines are limited in how much power they can draw before processor speed is automatically reduced. At any rate, since data center physical infrastructure is most often sized based on a high percentage of the nameplate ratings of the IT gear, this type of

variation in power is unlikely to cause capacity issues related to the physical infrastructure particularly when the percentage of virtualized servers is low.

A highly virtualized environment such as that characterized by a large cloud-based data center, however, could have larger load swings compared to a non-virtualized one. And, unless they are incredibly well-planned and managed, these could be large enough to potentially cause capacity issues or, at least, possibly violate policies related to capacity headroom.

Increasingly, managers are automating the creation and movement of VMs. It is this unique ability that helps make a virtualized data center more fault-tolerant. If a software fault occurs within a given VM or a physical host server crashes, other machines can quickly recover the workload with a minimal amount of latency for the user. Automated VM creation and movement is also what enables much of the compute power scalability in cloud computing. **Ironically, however, this rapid and sudden movement of VMs can also expose IT workloads to power and cooling problems that may exist which then put the loads at risk.**

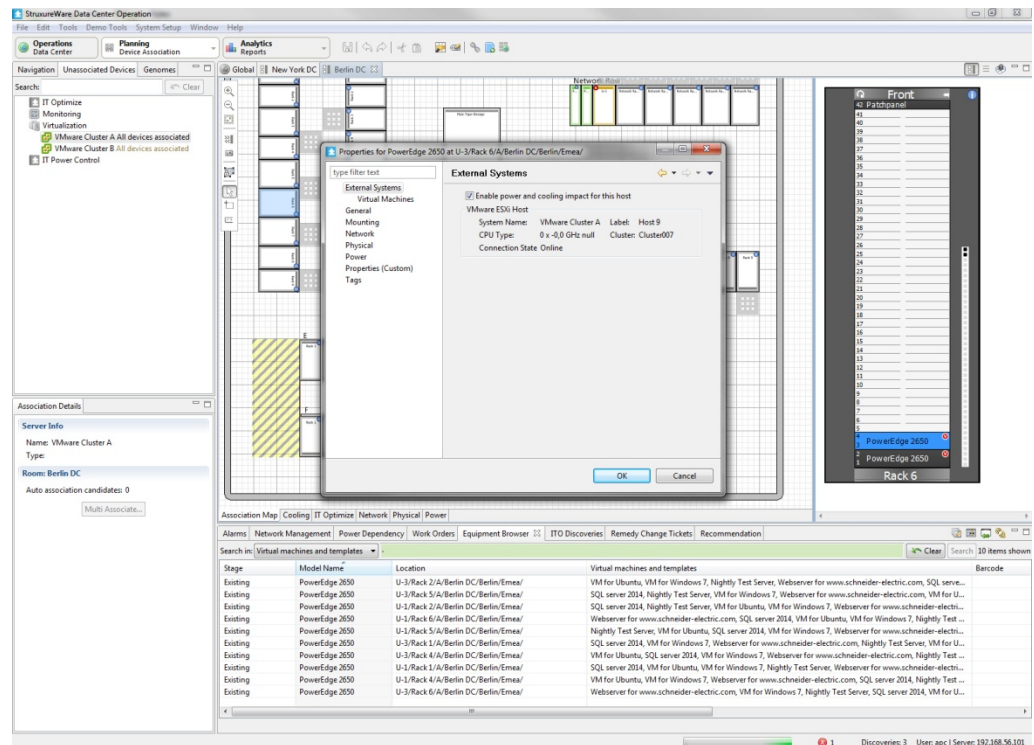## DCIM and VM manager integration to automate safer VM placement

Data center infrastructure management (DCIM) software can monitor and report on the health and capacity status of the power and cooling systems. This software can also be used to keep track of all the various relationships between the IT gear and the physical infrastructure. Knowing such things as which servers, physical and virtual, are installed in a given rack along with knowing which power path and cooling system it is associated with should be required knowledge for good VM management.   This knowledge is important because without it, it is virtually impossible to be sure virtual machines are being created in or moved to a host with adequate and healthy power and cooling resources.

Relying on manual human intervention to digest and act on all the information provided by DCIM software can quickly become an inadequate way to manage capacity, considering the many demands already placed on data center managers. The need for manual intervention introduces the risk of human error, a leading cause of downtime. Human error, in this case, is likely to take the form of IT load changes without accounting for the status and availability of power and cooling at a given location. Automating both the monitoring of DCIM information (available rack space, power, and cooling capacity and health) and the implementation of suggested actions greatly reduces the risk.

Data Center Infrastructure Management (DCIM) software is available that can provide this real-time, automated management. An example of this is shown in **Figure 7**. The two-way communication between the VM manager and DCIM software and the automation of action that result from this integration is what ensures physical servers and storage arrays receive the right power and cooling where and when needed.

A VM is created or moved to a different physical server typically because there are not enough processor, memory, or storage resources available at a given moment and location. But an effective management system can directly cause VMs to move based also on real-time, physical infrastructure capacity and health at the rack level. When DCIM software is integrated with the VM manager, VMs can be safely and automatically moved to areas known to have sufficient power and cooling capacity to handle the additional load. Conversely, VMs can be moved away from racks that develop power or cooling problems. For example, if there's a power outage at a rack, a cooling fan stops working or there is a sudden loss of power redundancy, the VM manager can be notified of the event and the "at risk" VMs can be moved to a safe and "healthy" rack elsewhere in the data center. All of this happens auto-matically in real time without staff intervention. DCIM software integration with a VM manager is a key capability for ensuring that virtual loads and their physical hosts are protected. In turn, service levels will be more easily maintained and staff will be freed from having to spend as much time physically monitoring the power and cooling infrastructure.

This integration becomes even more critical as power and cooling capacities are reduced or rightsized to fit a newly virtualized or consolidated data center. The less "head room" or excess capacity that exists, the less margin for error there is for placing virtual machines. **Maintaining a highly efficient, leanly provisioned data center in an environment characterized by frequent and sudden load shifting requires a management system that works automatically in real time with the VM manager.**

Also, it should not be forgotten that IT policies related to VM management need to be constructed so that power and cooling systems are considered. This must occur in order for the DCIM software integration with the VM manager to work as described above. Policies should set thresholds and limits for what is acceptable for a given application or VM in terms of power and cooling capacity, health, and redundancy.

Lowered redundancy requirements

The reduced need for power and cooling capacity is a commonly known benefit of IT virtualization as described in the prior section, *Reduced IT load can worsen PUE*. A lesser known benefit, however, is that IT virtualization can lead to a *reduced need for redundancy* in the physical infrastructure. Relying more on the fault-tolerance of well-managed virtual machines (and the applications that run on them), and less on higher levels of infrastructure redundancy could simplify design, lower capital costs, and save space for other needed systems or for future IT growth.

A highly virtualized environment is similar to a RAID array in that it is fault tolerant and highly recoverable. Workloads, entire virtual machines, and virtualized storage resources can be automatically and instantly relocated to safe areas of the network when problems arise. This shifting of resources to maintain uninterrupted service occurs at a level that is essentially transparent to the end user. However, depending on the quality of IT implementation and the level of integration of the virtual machine (VM) manager software, the end user may experience momentary latency issues while this migration takes place. But, generally speaking, service levels can be maintained in a highly virtualized environment even when some servers or racks become unavailable.

Given this fault tolerance, there may be a reduced need for a highly redundant (i.e., 2N or 2N+1) power and cooling system in a highly virtualized data center. If, for example, the failure of a particular UPS does not result in business disruption, it may not be necessary to have a backup, redundant UPS system for the one that just failed. Those planning to build a new data center with "2N" redundant power and cooling systems, perhaps could consider building an N+1 data center instead. This would obviously significantly reduce capital costs and simplify the design of the infrastructure. It's the fault tolerance of a highly virtualized network that allows organizations to consider this reduced infrastructure redundancy as a real option now. Before making these types of decisions, of course, IT and Facilities management should always fully consider the possible impacts to business continuity if the physical infrastructure system or component being considered fails or becomes unavailable. This means IT management systems and policies should be reviewed and monitored to ensure they are capable of providing the level of service and fault tolerance that permits having less redundancy in the physical infrastructure. This matching of physical infrastructure redundancy to the fault-tolerant nature of a virtualized IT environment is another form of the rightsizing described earlier. Rightsizing in this way can further reduce energy consumption, capital costs, and fixed losses all while improving the data center PUE.

Link to resource
White Paper 75

Comparing UPS System Design Configurations

For more information about UPS redundancy, see White Paper 75, *Comparing UPS System Design Configurations.*

To learn more about the possible impact of physical infrastructure design on data center capital costs, see the *Data Center Capital Cost* TradeOff Tool™. A link is provided in the *Resources* section at the end of this paper.

Conclusion

Virtualizing a data center's IT resources can have certain consequences related to the physical infrastructure. If these impacts and consequences are ignored, the broad benefits of virtualization and cloud computing can be limited or compromised, and in some cases, severely so. Areas of high density can develop after server consolidation takes place which may result in hot spots that can then lead to hardware failure. Various methods exist to ensure the cooling system has the means and capacity to reliably cool high density equipment. PUE can significantly worsen after consolidation occurs. By optimizing the power and cooling systems to better match this now reduced IT load, PUE can be significantly improved. This optimization is made much easier if scalable and modular systems are used. Dynamic IT loads which can change *automatically* in both time and location may unintentionally be put at risk if power and cooling status is not first considered at an individual rack level. Careful planning and on-going management is required to ensure VMs are only placed where healthy power and cooling exists. By constructing sound VM policies and by integrating DCIM software with the VM manager, this on-going management can be automated. Finally, the high level of fault tolerance that is possible with today's VM manager software makes it possible to employ a less redundant power and cooling infrastructure. Such a strategy can save time, space, energy and significantly lower capital costs. Implementing the solutions described in this paper will keep a highly virtualized data center running with greater reliability, efficiency, and with expanded flexibility to meet highly dynamic compute power demand.

✎ About the authors

Suzanne Niles is a Senior Research Analyst with Schneider Electric's Data Center Science Center. She studied mathematics at Wellesley College before receiving her Bachelor's degree in computer science from MIT, with a thesis on handwritten character recognition. She has been educating diverse audiences for over 30 years using a variety of media from software manuals to photography and children's songs.

Patrick Donovan is a Senior Research Analyst with Schneider Electric's Data Center Science Center. He has over 16 years of experience developing and supporting critical power and cooling systems for Schneider Electric's IT Business unit including several award-winning power protection, efficiency and availability solutions.

Resources

Click on icon to link to resource

The Advantages of Row and Rack-Oriented Cooling Architectures for Data Centers
White Paper 130

Deploying High-Density Zones in a Low-Density Data Center
White Paper 134

Power and Cooling Capacity Management for Data Centers
White Paper 150

Electrical Efficiency Modeling for Data Centers
White Paper 113

Comparing UPS System Design Configurations
White Paper 75

Browse all white papers
whitepapers.apc.com

Virtualization Energy Cost Calculator
TradeOff Tool 9

Data Center Capital Cost Calculator
TradeOff Tool 4

Browse all TradeOff Tools™
tools.apc.com

Contact us

For feedback and comments about the content of this white paper:

Data Center Science Center
DCSC@Schneider-Electric.com

If you are a customer and have questions specific to your data center project:

Contact your **Schneider Electric** representative at
www.apc.com/support/contact/index.cfm

Appendix

This case study was performed using the APC TradeOff Tool™ Virtualization Energy Cost Calculator. This hypothetical 1 MW data center is assumed to be 70% loaded with no redundancy in power and cooling with an electric bill of over $1,400,000 (**Figure 8**).

The data center is virtualized with a 10-to-1 decrease in server power. Because of the load reduction from virtualizing, the power and cooling infrastructure is now underloaded, and operates at reduced infrastructure efficiency (PUE 2.25) due to fixed losses, as described in this paper. At this stage, the total electrical bill would be reduced by only 17%. However, the PUE would further decrease to 1.63 and the electricity savings would reach 40% if the UPS, PDUs, and cooling units were also rightsized and air containment (e.g. blanking panels installed) methods employed. It should be noted that compared to the load after virtualization prior to optimizing the infrastructure, the overall data center load on the physical infrastructure would increase due to the rightsized data center capacity. And this would mean that the rated capacity of the overall data center would decrease given that the physical infrastructure had been rightsized to the lower IT load.
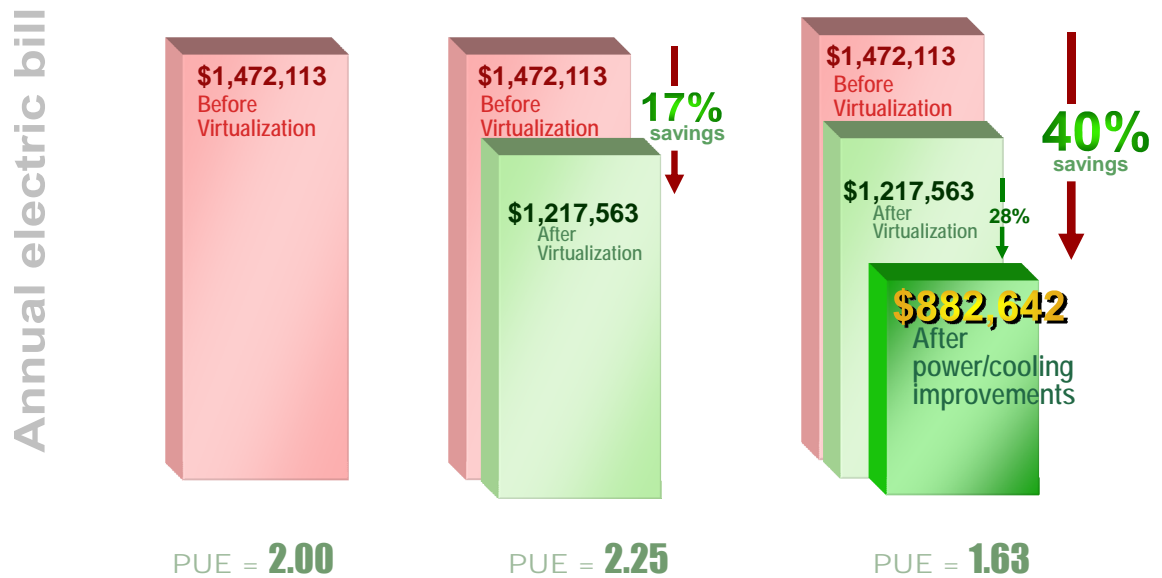
Figure 8

Case study showing effect of virtualization with and without power/cooling improvements

a. Before virtualization

Data center capacity: 1,000 kW
Total IT load: 700Kw
Data center loading: **70%**
Server load: 490 kW (70% of IT load)

Room-based cooling
Hot aisle/cold aisle layout
Uncoordinated CRACs
UPS: Traditional, 89% efficient at full load

18" raised floor with 6" cable obstruction
Scattered placement of perforated tiles
No blanking panels

b. After virtualization
   Server consolidation only

Data center capacity: 1,000kW
Total IT load: 515 kW
Data center loading: **52%**
10-to-1 server reduction
100% of servers were virtualized
Server load: 305 kW (59% of IT load)

No change to DCPI elements

c. After virtualization
   PLUS power/cooling improvements and rightsizing implemented
Data center capacity: 515 kW
Total IT load: 515 kW
Data center loading: **100%**

UPS: High efficiency, 96% efficient at full load
Row-based cooling (no containment)
Blanking panels added
CRAC, CRAH, UPS and PDUs rightsized to load

NOTE: 90% of these total energy savings were gained through optimizing the existing system (high efficiency UPS, installing row-based cooling and blanking panels). The remaining 10% came from rightsizing the CRACs, CRAHs, UPSs and PDUs.