Cisco Energy Efficient Data Center Solutions and Best Practices

Introduction

It has taken more than 30 years for the data center to evolve to a point of modularity that is conducive to alignment of efficiencies in use and design across disparate infrastructures. The modern environmental, or "green", movement has also taken 30 years to evolve to a point where the economics of sustainable practices are well-understood. Traditionally, environmentalism has been perceived to be at odds with economic prosperity. Due in part to advances in information technology, communications, and computer-aided design (CAD) and planning, this mindset is being challenged in the IT industry today. IT professionals have been asked to do more with less for some time now and therefore are quite familiar with achieving efficiencies in applications and data center operations.

This document discusses how IT professionals can better design power and cooling efficiencies in their data centers through the strategic application of Cisco® and partner technologies. It seeks to empower IT professionals to work more effectively with their facilities counterparts. Because the entire industry is in the planning phase regarding green data center operations, the intent of this document is to share best-practices solutions that Cisco has in place today. Cisco will continue to provide leadership for its customers and partners on the principles of green networking.

Evolving Towards a Green Data Center

Because there is no global, industry-wide recognized standard today that defines a green data center, each organization needs to define what green means to it. For the sake of this document, a data center is generally considered a purpose-built facility or portion thereof that houses a business's mission-critical applications. A green connotation supposes that the facility has either been retrofitted or initially designed to mitigate its power consumption through the strategic application of available technologies. The long-term goal of the green data center operation is to achieve carbon neutrality. This document does not specifically focus on green building, which covers aspects outside the four walls of a data center.

Whatever a data center manager's definition of green, a moral, financial, or capacity motivation to pursuing this agenda can be assumed. For a business operation to demonstrate success in implementing a green strategy, the strategy must be measurable. A variety of measurement methodology options are available today. Many of these options may be too costly in terms of time and resources to implement when the scope of internal integration is accounted for. To IT, knowledge is power, but the collection and interpretation of the vast quantities of data in a data center must be simple, intuitive, and reliable.

Recent times have seen a surge in new technologies that have driven up power capacity and density requirements in IT, mainly in server and storage-related operations. The data center network has not traditionally been a major contributor to the power problems faced in today's data center, mainly because of the relatively low consumption of power by the network compared to the overall power consumed in a data center. Most estimates of IT infrastructure consumption range from 8 to 12 percent for the network.

Historically, power consumption requirements by the network have been relatively low because of the comparatively low-level of computing functions in switches and routers. As the network has evolved to include higher levels of intelligence, its corresponding power requirements have grown and will continue to do so. This correlation between function, intelligence, feature sets and corresponding power requirements is a crucial concept. Ultimately, the end user must decide how to balance power consumption with service levels for voice, video, and data.

The Blade Era Meets the Green Era

Much attention has focused on blade form factors in server platforms in recent years. Although true blade servers do present density and capacity concerns for older data centers, these concerns are distinct from concerns about the efficiency of electrical consumption.

Where the concept of green is concerned, efficiency is the focus. More effective use of existing power supply corresponds to fewer emissions resulting from commercial power generation. From this perspective, blade designs are good. Blade form factors per work unit performed consume power and cooling more efficiently than do standalone 1-, 2-, and 4-rack unit (RU) servers.

Because data center efficiency is in part a cumulative calculation of how disparate IT infrastructure interoperates with supporting facilities, blade capacity and density requirements must be aligned with facility design to achieve maximum efficiency benefits. In most cases, this requirement often necessitates building a new data center to support higher densities, mainly because of the cost of upgrading the existing facilities infrastructure to support higher densities.

Gartner Research recently analyzed the cost of retrofitting an existing 800-kilowatt (kW) data center designed to support rack densities from 2.5 kW per rack to 20 kW per rack through the introduction of blade servers. The effect of these servers was based on replacing one-third of the capacity of the data center with high density blade configurations. The study found that the annual estimated operating expense rose from US$800,000 to US$4.6 million (Table 1).

Table 1. Gartner Research 2006 Analysis: Affording the Next-Generation Data Center

	Traditional Server	High-Density Server
Power Per Rack	2-3kW per Rack	~20kW per Rack
Power per Floor Space	30-40W per Square Foot	700-800W per Square Foot
Cooling Needs: Chilled Airflow	200-300 Cubic Feet per Minute (CFM)	3000 CFM
Total Annual Operating Expense = US$800,000		
Traditional Data Center Designed to Accommodate 2-3kW per Rack		
Total Annual Operating Expense = US$4.6 Million		
Implementing High-Density Solutions in Older Data Centers is Often Too Expensive		

In this case, the deciding factor is whether the cost of the supporting equipment outweighs the capital and operating costs of the IT equipment it supports; this misalignment is the cost of not aligning facilities design with the design of the IT infrastructure. However, most IT professionals who are responsible for purchasing IT equipment have little insight into the facilities that support their infrastructures.

This misalignment and its corresponding capacity and density problems occur in part because the facilities that support a given operation have much longer refresh cycles than do the IT equipment they support. Over the past decade, servers and storage arrays have typically been refreshed every 2 to 5 years and large switches and routers every 5 to 7 years. Data center facilities are often designed with a 10- to 20-year lifecycle. Consequently, the pace of technology adoption has traditionally been much slower for facilities infrastructures than for IT.

This mismatch between facilities and IT has been problematic for many end users in both industries. Today's digital consumer continues to ask for better processing performance, more storage space, improved access, and lower latency, causing IT vendors, under constant cost pressures from shrinking IT budgets, to focus increasingly on optimizing their solutions. However, innovation has limits where the physics of power and cooling mechanics are concerned. Therefore, it is important to understand that IT infrastructure performance correlates directly to the size, weight, power, and cooling requirements of a given solution.

With the introduction of blade and service module form factors, the IT industry realized it could do much more with less space. The capacity problems that have arisen have led many in both the facilities and IT industries to look more closely at optimizing existing data centers to accommodate desired performance levels, often to gain some time to plan a new facility. It has also led vendors in the IT space to look more closely at designing products with facilities considerations in mind. Further, this focus on power and cooling has brought today's data center to the forefront of the green agenda because it is clear that green sense makes economic and operational sense in this IT environment.

Difficult questions need to be addressed internally for a given organization and operation to develop a green agenda. Gartner Research recently conducted a study that highlights the need for both vendors and users to be more aware of their effects on the environment. "Few IT management teams are aware of their enterprise's corporate social responsibility and environmental policies, and they have not mapped out the implications for their own activities," says Simon Mingay, research vice president at Gartner Research. "They need to decide whether to take a proactive response, a measured response following the market and legislation, or a passive approach that just meets legal requirements. The roles, responsibilities and programs will be very different for each."

Regulatory Compliance

Although no global or broadly accepted standards exist today for government bodies to regulate data center efficiency, there is anecdotal evidence that they are being considered. With the passage of the U.S. House of Representatives Resolution 5646 in July 2006, the U.S. Environmental Protection Agency's Energy Star program is now researching data center energy consumption, industry measures for server efficiency, and energy-saving technologies.

The recently formed Regional Green House Gas Initiative (RGGI), led by nine U.S. states, is a state government-level program that seeks to provide flexibility in the way that emissions from commercial operations are managed. This concept of self-regulation in an open-market framework through the trading of carbon credits is similar to the concept behind many programs that exist in the European Union.

In Europe and Japan, numerous regulations that address reduction, reuse, and recycling in IT-related operations have been developed. In Europe, the Restriction of Hazardous Substances Directive is one of the most mature programs with implications for IT manufacturers. This design of this program is similar to regulations developed by Japan's Ministry of Tourism, Trade, and Industry (MTTI).

Myriad public and private efforts are now directed at addressing the broad range of concerns related to global climate change. Clear leadership is emerging within the IT industry, which is positioned very well to maximize the benefits of existing technology while investing in future green technologies. To this end, venture capital spending in the Silicon Valley directed at green technologies doubled from 2005 to 2006.

One of the biggest challenge facing legislators, corporations, and management agencies today is the fact that there is no simple, accurate way to measure and monitor power consumption and its corresponding emissions in real time. This lack is particularly problematic for regulation, and until a comprehensive management system is developed and implemented, green behavior will be a largely voluntary exercise.

Measurement Considerations

According to direct customer feedback from Cisco and its partners, most data center managers do not know how efficiently their data centers are operating today. In some cases, the facilities department may have insight into power efficiencies between the main feeds to the data center and the main feeds to a campus or building.

The main reason for this lack of knowledge is that today measurement is a manual process, and results can vary widely within several hours of performing a one-time calculation. However, it is important to establish an annual baseline, even if the measurement is performed manually, if a data center manager is to develop policies and processes to implement best practices for data center power consumption.

A data center-wide efficiency calculation can be approached in several ways. If the calculation is to be performed in-house, it must include the facilities department to address all equipment that supports (even fractionally) the data center operation. Many IT professionals opt to obtain expertise through professional service organizations and consulting engineers because most IT operations do not have the necessary skill sets.

Several metrics exist today that can help determine how efficient an operation is. These metrics apply differently to different types of systems: for example, facilities, network, server, and storage systems. For instance, Cisco uses a measure of power per work unit performed instead of a measure of power per port because the latter approach does not account for use cases: the availability, power capacity, and density profile of mail, file, and print services will be very different from that of mission-critical Web and security services. Furthermore, Cisco recognizes that just a measure of the network is not indicative of the whole data center operation. This reason is one of several reasons why Cisco has joined The Green Grid (http://www.thegreengrid.org), which focuses on developing data center-wide metrics for power efficiency. The power usage effectiveness (PUE) and data center efficiency (DCE) metrics detailed in the document "The Green Grid Metrics: Describing Data Center Power Efficiency" are ways to start addressing this challenge. Additionally, an IT professional can calculate how differences in efficiency will affect the cost of data center power. A good place to start this analysis is with air conditioning. Consider the operating cost of a 1000-kW computer room air conditioning (CRAC) system with a 60-percent efficiency at US$0.10 (U.S. average for power) per kWh for one year at a 50-percent load level:

```
Operating cost = 1000 kW x (1/0.60 efficiency at load level) x 8760
hours x US$0.10 x Load level 50%.
PUE = Total facility power / IT equipment power
DCE = IT equipment power / Total facility power
```

The first option, which is also the simplest, involves adding a few columns to an existing asset inventory list. The data-collection process involved for any option is similar and involves inventorying each component in the data center that consumes power, including everything from lighting to servers to CRAC systems. Any component that consumes power has certain efficiency losses based on the manufacturer's specification. It is important to develop a loss profile based on each infrastructure segment. This step is necessary in aggregating box measurements to systems measurements and ultimately achieving a data center measurement.

From a risk standpoint, if one assumes that regulations regarding data center efficiencies will ultimately be implemented, the data center manager should begin to establish baselines to lessen the operational effect of these regulations when they arrive.

Power Densities and Facilities Design

Many of the problems being experienced today are exacerbated by a lack of consistent and common communications between facilities and IT. Professionals in each industry have different operation agendas, budgets, and management platforms. Communication between these groups can be difficult to manage and is often neglected.

The IT group needs to attain sufficient baseline understanding of power consumption to effectively communicate with the facilities group as one of the first steps in assessing IT infrastructure to support a given data center project. Ensuring that this step is addressed before any other IT prioritization often saves wasted project planning cycles. The advent of blade technology has forced facilities and IT groups to examine the short-to-midterm misalignment between the densities specified by IT manufacturers and the densities that a facility is designed to support.

A best practice in planning the density and capacity of a data center is to move away from the common watts-per-square-foot measurement. Any data center manager can attest that a data center is almost never deployed uniformly after the first year. Adhering to a watt density per rack helps give the data center manager more control over the standard operating environment (SOE) per rack. From a planning and deployment standpoint, this variable is a much more manageable building block that can be used in a zone-based approach.

For the purposes of this document, a data center zone is defined as a logical and physical group of assets that supports certain applications or a subset thereof. The concept of zoning or clustering is not new to IT. However, some changes in the past few years in data center facilities allow this concept to be better adopted at the physical layer. A well-planned zoning of the data center across facilities, network, server, and storage allows managers to more easily align themes such as criticality, density, capacity, efficiency, and scalability across these infrastructure segments.

The practice of zoning is particularly important in a virtualized environment. For example, a data center may have 100 servers that support three very different applications. Each of these applications has different criticality requirements, but all may be running in any microenvironment within the data center at any given time. Any of these microenvironments may be peaking over the specified temperature thresholds, regularly causing a threat to the availability of servers. This incremental threat of hardware failure can be quantified using the Arrhenius equation, which states that for every 10 degrees Celsius that a piece of equipment operates nominally above specification, the lifespan of that component is essentially halved.

Power density and facility design requires a balance between many factors, but with the right planning, the defining elements of capacity, density, criticality, efficiency, and scalability can be aligned through the infrastructure and applications. Some basic best practices can be adopted to begin planning the design of a highly efficient data center operation:

- Adopt the rack as the basic building block for data center density. Watts, ports, I/O operations, processors, and disks per rack are much more accurate measures than watts per-square-foot or -meter measures because data centers are rarely deployed uniformly.
- Set internal standards for the alignment of SOEs, from facilities to applications, including racks, rows, zones, and rooms.
- Define SOEs to stipulate that capacity, density, criticality, and scalability must be aligned for each infrastructure and application tier.

Cooling

Typically, the largest consumer of power and the most inefficient system in the data center is CRAC. Fully half the power in the data center goes to cooling, mainly because of the burden factor of air conditioning, which is the ratio between production power and cooling power. This ratio typically ranges from 1.8 to 2.5, depending on many factors; if a server requires 100 watts of production power, a CRAC unit would require between 180 and 250 watts of power to cool the server. The factors that contribute to the burden factor of a CRAC unit are mainly the operational efficiency and air distribution design of the unit, or how efficiently the air can reach the server it is cooling.

Examining air flow in the data center typically uncovers many simple fixes that can increase the efficiency of the entire cooling system. Cisco and its partners help customers implement many best practices. In some cases, these fixes are as simple as installing blanking panels in open rack space, rearranging perforated tiles, and installing hot-air and cold-air partitioning. In other cases, they may be more complex retrofit projects such as migration of cabling out of the subfloor plenum to eliminate constraints to air flow. The best time to design efficiency into an air-distribution strategy is when building a new data center that can accommodate densities in line with current IT infrastructure projections. Using watt and cfm-per-rack specifications and implementing power, temperature, and humidity monitoring over the network will help efficiency measurement move forward.

When looking for areas on which to focus efficiency efforts, consider demand fighting between CRAC units, which is the largest contributing factor to inefficiency. Demand fighting refers to the common scenario in which unsynchronized air conditioning units "fight" each other over temperature, humidity, and air flow supply. Bringing CRAC units into an IP-based network and synchronizing them can often double the efficiency of the biggest power consumer in the data center by mitigating or even eliminating demand fighting. A user can also monitor the delta temperature value between the inlet and exhaust of IT assets to help ensure that the delta value is as close to 9 degrees Celsius as possible.

Use of an IP-based network with open protocols is also crucial to measuring the efficiency of a cooling system. This measurement can be accomplished by providing a network connection to each cooling component and aggregating that to a server-based management platform such as the Dynamic Host Configuration Protocol (DHCP). For older CRAC units that do not ship with an

Ethernet port, a simple protocol converter can be installed that converts building management system (BMS) protocols such as Modbus to the Simple Network Management Protocol (SNMP) or an equivalent protocol. Some manual naming of the converted data will be required upon initial setup in this scenario, but the efficiency gains justify the resource investment.

Rack Environment

For IT professionals, the rack should be the primary building block for power and cooling planning. Starting with the rack gives the planner a higher level of control when considering the peripheral supporting infrastructure and adjacent effects of IT infrastructure.

Watts can be adopted as the unit of measurement for both the power and cooling of a given rack, simplifying the higher-level planning for rack deployment. For air flow within the rack, a cfm measure should be employed because air distribution is as critical a factor as air supply. One cfm equals about 28.31 liters per minute.

Using a watts-per-rack measurement standard, the data center manager can plan more effectively after the initial infrastructure deployment in the data center. In contrast to a watts-per-square-foot measure, which is much more difficult to quantify after the main planning phase ends, a watts-per-rack standard lets a manager set standards and policies for incremental additions of IT infrastructure and better plan how to scale the facilities infrastructure to meet IT demand.

Additional considerations within the rack that can have a profound effect on the densities that can be stipulated are in-rack structured cabling, air flow direction resulting from fans within the IT equipment, efficiency losses of the IT equipment, open rack units, and intake and exhaust juxtaposition. These factors should all be considered when designing SOE guidelines.

IP Management and Building Management Systems

Historically, facilities infrastructure, including uninterruptible power supplies (UPSs), computer room air handlers (CRAHs), CRAC, temperature and humidity monitors, fire suppression, physical security, power distribution units (PDUs), and branch circuits, are still not monitored in an IP-based network. This lack provides the IT professional with significant challenges in planning incremental additions and leads to a one-way communications model, from the IT group to the facilities group.

Significant advances in the management capabilities of facilities infrastructure in the past decade are allowing IT professionals to begin to monitor, measure, and manage the physical layer through an IP network in parallel with facilities professionals. The importance of aligning facilities infrastructure with IT infrastructure has already been mentioned. Implementing a network-based monitoring strategy allows the data center manager to measure progress toward goals to improve power efficiency.

For its internal IT department, Cisco has implemented a temperature sensor network (TSN), which provides the first level of analysis for correlating the demand of IT equipment and the supply of facilities infrastructure. Many efficiency gains can be identified in the relationship between facilities supply and IT demand. Providing cooling on demand is always more efficient than overcooling, and the TSN helps ensure that overcooling and demand fighting do not occur. Proactive network monitoring delivers the data needed to make more informed decisions about how to achieve efficiencies in the system.

In summary, because data center efficiency losses are roughly 80-percent process and 20-percent infrastructure losses, IP-based network management can have the greatest effect on data center efficiency at the operations level. A focus on components alone will not have as significant an effect on the overall system.

Network Architectures

As stated in the previous section, data center efficiency must be quantified at the systems level and aggregated in a sitewide measurement. The position of the network is unique among the IT infrastructure segments. The network is the ubiquitous platform in the data center and is ideally suited to bridging the gap between facilities and IT.

When considering network architecture, a planner should not confine the scope to switches, routers, load balancers, and firewalls. Where possible planners should look to extend the touch of the network to gather as many relevant data sources as possible to monitor power efficiency. These sources include monitoring of power, temperature, and humidity; UPS, CRAC, CRAH, branch circuits, PDUs, and specialized sensors that are typically found inside or close to the main floor of the data center. Including these components is nondisruptive and cost-effective where connectivity is concerned because they can be included as part of the existing or normally planned structured cabling plant. For components such as gensets, chillers, and some transformers, some additional cabling might be necessary, but the cost of including these components are usually insignificant.

In addition to using the network as the physical “middleware” between IT and facilities, efficiency opportunities in network design to support different services are significant. If application delivery, security, access and identity, storage, and computing are considered as services, the network can have a profound effect on the cumulative efficiency of an operation. The following sections explore the implications of the right network architecture on power, cooling, and management. Furthermore, migration strategies and best practices are suggested.

The Network Effect

The network effect refers to the collapsing of services or portions thereof provided by standalone components into a more ubiquitous network-based platform. The precedence of the occurrence of this collapse is clear since the inception of switching and routing. Print, storage, and file services are obvious examples, but this trend continues to now cover enterprisewide application delivery, security, Web, identity, and access services. When analyzing the ratio of power used to service delivered, it is very important to consider the effects of the network across a broad range of disparate infrastructure and service levels as a first step. For the sake of illustrating the network effect, the primary focus is on service delivery as it relates to use of assets and efficiency in network design.

Numerous studies in the past two years have indicated that server and storage utilization rates range from 20 to 40 percent. Of course this level of inefficiency in asset utilization has a corollary effect on power usage in a given operation. Utilization can be increased in many ways that are process-oriented - the larger proportion of the variable - but taking advantage of technology in the network today can have a profound complementary effect. The next section details the effect of the network on power and cooling resulting from network-based virtualization and the corresponding higher utilization of infrastructure.

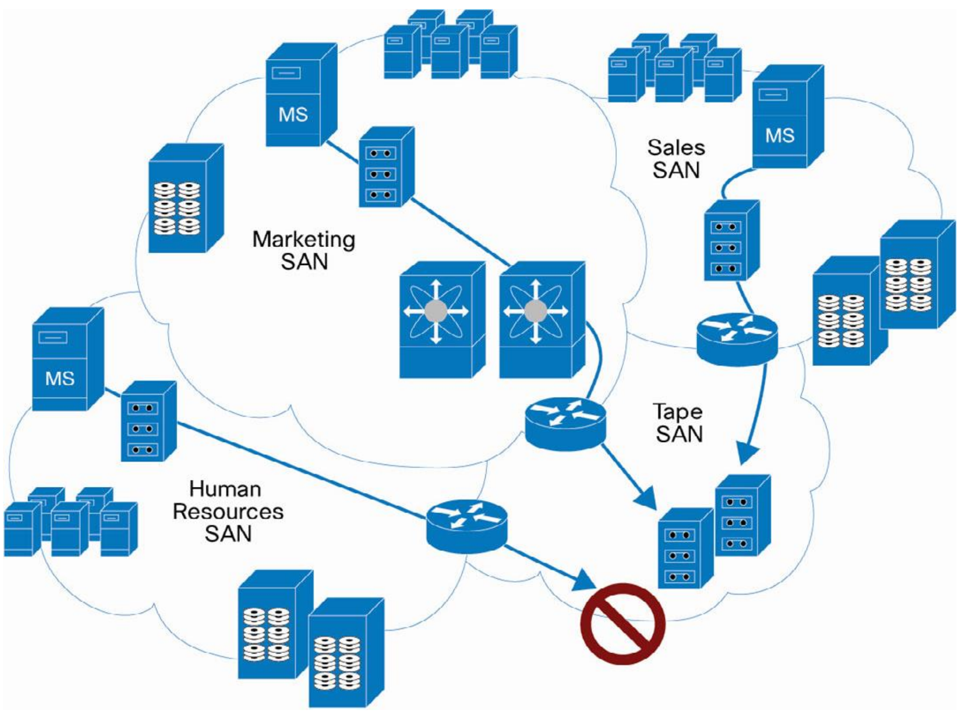
Power and Cooling Implications

An operative assumption for designing the infrastructure for a data center would be that the fewer power supplies in a data center, the fewer points of conversion loss. Anytime power traverses a power supply in the data center, whether it is a switch mode power supply (SMPS) or large UPS system, it looses some of its capacity in the form of heat. This loss can be generally quantified based on the manufacturer's power supply efficiency specifications. An example can be found in most of Cisco's larger switching platforms, which have supplies that operate at 90-percent efficiency. If one of these switches requires 1000 watts to operate, it can be assumed that 100 watts will be lost through conversion as heat - a loss that in turn carries a burden factor on cooling.

If this supposition is carried throughout the planning cycles, the user is left with an efficiency comparison of not just heterogeneous systems but systems across all IT infrastructures. Now that many infrastructure systems (that is, server, storage, and network) in the data center are achieving parity in the services they can support or deliver, this consideration is an important one to make when deciding where services should reside.

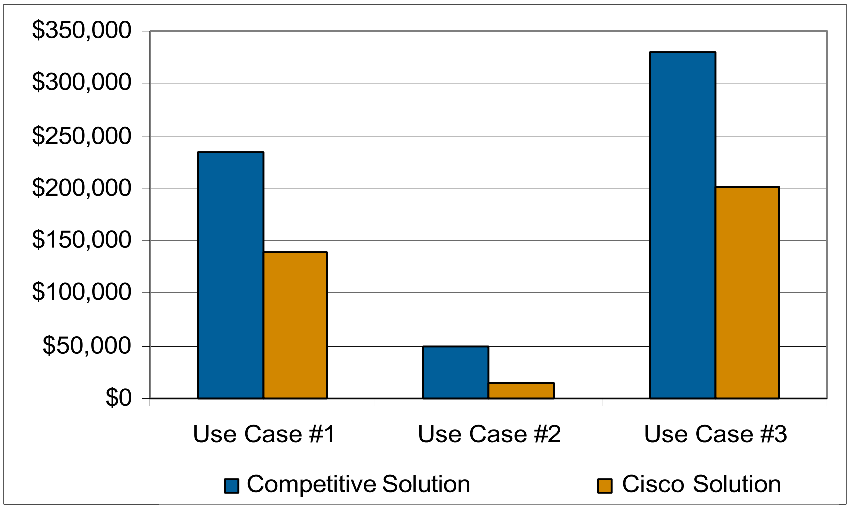
If fewer power supplies and superior or similar service levels are assumed, virtualization of computing, storage, and networked resources on a common platform becomes even more attractive. Virtualization specifically has the potential for wholesale effect by increasing utilization rates of assets, thereby allowing more from less (Figure 1).

Figure 1. Using a Common Physical Fabric to Increase Storage Utilization Through Inter-VSAN Routing



Another benefit of virtualized services over the network relates to the collapsing of functions found in standalone appliances in the network. As more intelligence and feature sets are brought into a service module form factor, the less power per work unit performed is required - again, doing more with less (Figure 2).

Figure 2. Use Case Analysis of Security and Application Delivery Services with a Service Module Architecture versus Appliance-Based



Source: ESG White Paper – Building Power Efficient Solutions with Cisco MDS 9000 Directors

Use case 1: Isolated storage traffic for three BU's business units sharing a common backup.

Use case 2: Combined Mainframe and Open Systems storage network connectivity in the same data center.

Use case 3: Extended storage for remote replication and tape vaulting between data centers.

Optimization Points

Exploring the network as a platform for services introduces many new opportunities to increase capacity, accelerate efficiencies, and slow power capacity growth. At the highest level, examining how applications are delivered and stored with power in mind and at the enterprise level are questions worth asking. However, this process may involve a wholesale reevaluation of how the current IT architecture supports the business.

If an organization is serious about stabilizing or reducing emissions growth through the strategic application of IT, then it must consider all technologies. In the case of the network, given its ability to touch anything that consumes power, fixed or mobile, by definition it will identify more points of optimization than any other system.

Examination of most business operations indicates that the 80/20 rule applies to power consumption efficiency. Although components that have low efficiency still exist, roughly 80 percent of the losses are from process and 20 percent from product design. This fact leads to the conclusion that to better plan processes planners need the right data to justify a change to the present system. The network is a tool to help accomplish this justification.

Many management interfaces that assist in the analysis of data gathered through the network are available today. Cisco and its partners offer solutions that give users with the right purview to identify focus areas for service optimization. A good way to start this analysis is by examining how a given operation delivers an application with the understanding that every time a packet is routed, switched, cracked, served, or stored, a power footprint is associated with that action. A power footprint corresponds to a carbon footprint. Using the right mix of management applications allows a user to identify the right focus areas to guide toward carbon neutrality.

Migration Strategies

After a baseline understanding of the fundamentals of data center power efficiency is internalized to IT and optimization points are identified, a migration strategy can be developed. For the sake of this paper, the implementation of this strategy would cover the operative efficiency of the data center and could include the physical migration of assets to different environments. The scope would, of course, be user-defined, but the following sections provide supporting data and best practices that can assist planners formulating strategies concerning the network. It is also important to note that the goal of migrating toward a carbon-neutral operation is a perpetual effort.

Users often supplement internal skill sets by engaging with professional service organizations, particularly when establishing an efficiency baseline for the data center operation, which can be a complex and manually intensive exercise. However, establishing a baseline is necessary in order to migrate to and improve the operative efficiency of the data center.

As mentioned earlier, calculating efficiency includes calculating total power consumed per system. However, this infrastructure assessment needs to be aligned to the services the data center offers. For example, of the systems that are calculated, what portions of the systems support security, access, application delivery, and so on? This analysis provides a power profile per service or application, allowing for a before and after analysis.

When a service power profile analysis is complete, more accurate decisions can be made for changes in the infrastructure architecture. The analysis should include facilities as well as IT infrastructure; then a sound migration strategy can be developed. This migration is from lower efficiency and utilization to higher, and it may or may not include a physical relocation of assets.

Network Monitoring and Management

One of the main reasons utilization rates and efficiencies have traditionally been low in data center operations is the lack of insight a user has to power allocation and corresponding efficiency. However, numerous developments in this space have occurred in the past five years that have given the IT professional more control of power consumption in the data center.

Recent years have seen significant changes in the facilities infrastructure that allow for the IP-based management of components such as UPS, CRAC, branch circuits, PDUs, and gensets. Many manufacturers ship these components today with Simple Network Management Protocol (SNMP) interfaces that can be brought onto the network. When the facilities equipment is migrated onto an IP-based network, a user can identify points for optimization and ultimately begin to correlate the relationship between IT equipment demand and facilities supply.

Cisco recently partnered with APC/MGE, a Cisco Technology Developer Partner, to develop a macro, which is incorporated into the Cisco VFrame Data Center (DC) 1.1 management and provisioning application. This macro bridges the gap between facilities and IT and empowers a manager to not only automate the provisioning of servers but also helps to determine the best data center zone in which to physically place the server. It accomplishes this by interfacing with an APC switched rack power distribution unite (PDU).

The switched rack PDU units provide critical power monitoring and outlet control data to the Cisco VFrame platform to be used when deploying server or storage configurations. This integration is an example of how APC/MGE and Cisco are leveraging joint technologies to enable IT and facilities professional better instrument their critical data center environments.

This gives Cisco VFrame DC a homogenous power management solution for servers, which is of particular benefit for infrastructure that does not have a native solution or for a user who wishes to standardize on one application for all server assets. Based on requirements defined for the logical elements of a service network, Cisco VFrame DC then allocates appropriate physical or virtualized resources and configures and activates the service network. In this way it enables server, storage, and network administrators to create a flexible infrastructure that can adapt to changing application requirements.

Figure 3. Cisco VFrame Data Center 1.1 Topology illustrating relationship between service templates, service networks, resource pools and facilities.

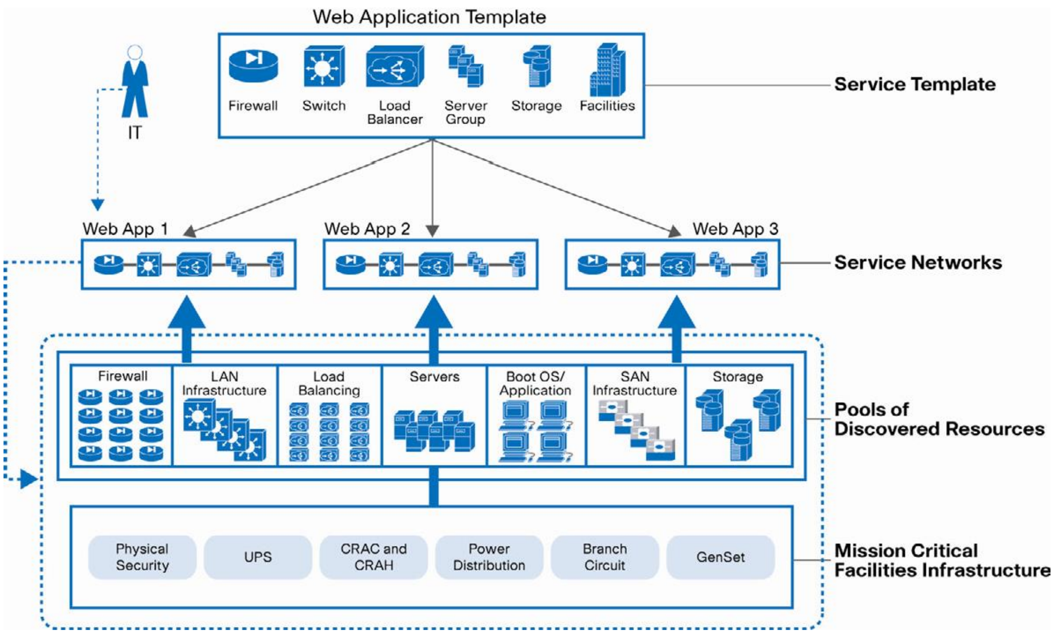
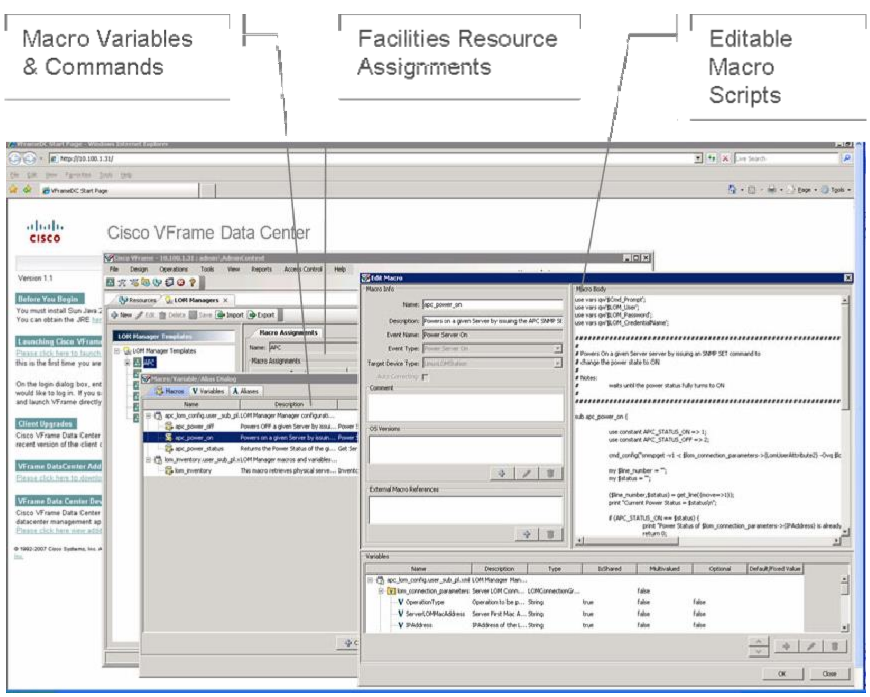


Figure 4. User Interface for Cisco VFrame DC 1.1 with APC/MGE Macro



Cisco partners offer complementary power management applications to enhance an IT-based power management strategy. IBM's PowerExecutive offers application support analysis based on administrative analysis of server power consumption profiles. This solution is available for selected IBM BladeCenter and System x servers and allows direct power monitoring through IBM Director.

Another Cisco partner, HP, has recently released a solution called Dynamic Smart Cooling (DSC). This technology helps users change data center energy costs from a fixed to a variable cost, increasing IT scaling headroom, and allows for higher efficiencies in the data center server architecture. This solution is designed to work with HP's C-class Bladesystem and ultimately correlates the demand of the servers with the supply of the CRAC. This latter solution feature set helps mitigate or eliminate demand fighting.

Finally, Cisco IOS® Software has offered management and control of switch and router chassis power characteristics for many years. A user can use Cisco IOS Software to determine if the power supplies in a given chassis are sized correctly. Ensuring a power supply is loaded to close to its maximum specification to load ensures higher operating efficiency of the component. Furthermore, a user can perform numerous controls through Cisco IOS Software, such as power cycling, temperature interrogation, and alerting.

Power Efficiency Considerations

Many vendors lead with messaging around power and cooling solutions that focus on cost but deliver this message to the IT professional. Although the IT department would certainly care about the company's overall financial position, power and cooling costs may not be a daily pressure for the IT department. Ultimately it is the facilities department that is responsible for the data center power bill.

Ensuring the facilities department is part of any conversation around power cooling with a vendor, integrator, or consulting engineering firm is crucial to understanding the overall value proposition of a given solution. To this end an IT or data center manager should consider a closer working relationship with the facilities department. If the facilities department is contracted and has little vested interest in innovation, users should consider supplementing these skills through third parties.

An organizational transformation trend that is emerging is the blending of skills between IT and facilities. For example, Cisco IT has two full-time facilities professionals, who focus on ensuring alignment in the planning, design, and operation of Cisco data centers.

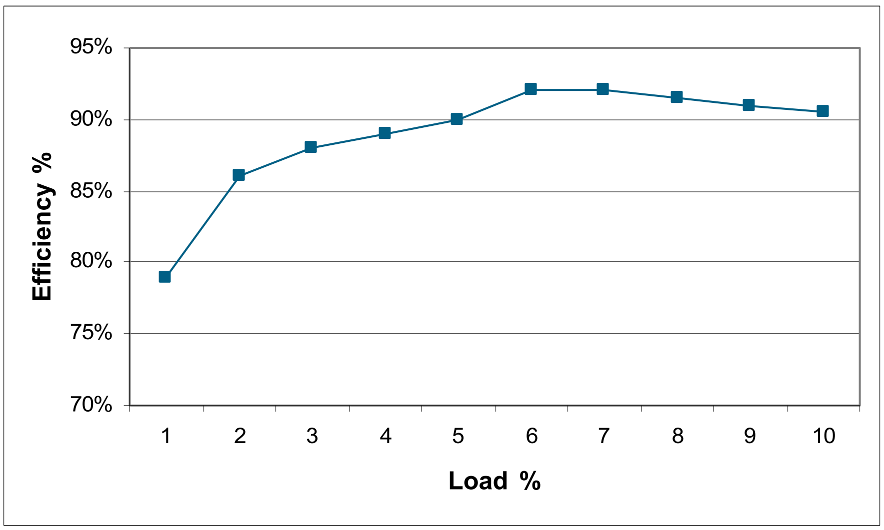
When IT and facilities departments collaborate, they can begin to engage at senior levels within their organization on funding power efficiency initiatives. This collaboration commonly involves the senior managers in both groups identifying solutions that reduce power costs and approaching the chief financial officer or equivalent with a co-developed plan. This plan should include an estimate on power savings for a given period of time with the stipulation that if the targets are achieved, IT and facilities split the savings. This methodology is a sound one for developing self-funding power efficiency projects.

Green Math

Although power and cooling can be analyzed in many ways, several established analysis methods can help a user analyze and benchmark data center efficiency. Efficiency considerations involve two main focus areas: the proportionate loss of power due to conversion loss in a given power-consuming component and a determination of the elements of all components that support a given service.

Consideration of power conversion losses should involve a review of the specified efficiency of a component. For example, Cisco AC power supplies for its larger switching platforms are 90-percent efficient when loaded at 85 percent or higher. If a data center manager wants to make an operation more energy efficient, a default standard in requests for pricing should include specification of highly efficient components. Furthermore, policies and management applications that ensure power supplies are correctly sized should be specified to ensure they are within 80 to 90 percent of load requirements. Figure 5 shows an example of an efficiency curve for the Cisco Catalyst® 6500 Series Switch.

Figure 5. Power Supply Efficiency Curve for Cisco Catalyst 6500 Series Switch



When designing efficiency into the services-oriented architecture (SOA), planners must consider the best practices discussed in the previous section. In addition to a simple aggregation and average of systems-level efficiency is the concept of assigning physical assets to logical services. Although this process can be very complex, it is a necessary first step to accurately compare and contrast service-oriented solutions with power as a consideration.

However, IT utilization rates should be examined more closely. Cisco offers very compelling solutions in this area that can increase the utilization rates of disparate server and storage infrastructures, effectively doing more with less. Cisco implemented Inter-VSAN Routing (IVR) in conjunction with a storage consolidation project that increased utilization on 2 Petabytes from 38 percent to 68 percent.

Users must determine what actions to take based on a utilization assessment, which may involve the removal or redeployment of low-value assets but more likely will help alleviate the effects of power utilization resulting from infrastructure sprawl. Users often can open a 12- to 24-month planning window for a data center retrofit, consolidation, or relocation. Table 2 shows an example of this analysis.

Table 2. Cisco Storage Consolidation with IVR: Power Implications Resulting from Higher Storage Utilization

SAN Consolidation: Power Considerations								
Before Cisco IVR		**Infrastructure:**	Total = Total archive units x Total 400W (redundant) supplies (800W)					
	Utilization	**Tape Archive**	**N + 1 SMPS (watts)**					
SAN 1	40%	10 x Generic	8,000					
SAN 2	50%	10 x Generic	8,000					
SAN 3	30%	10 x Generic	8,000					
SAN 4	40%	10 x Generic	8,000					
SAN 5	40%	10 x Generic	8,000	UPS	CRAC			
		Total power	**40,000**	**60,000**	**80,000**			
Storage Capacity	**Average Utilization**		**80%**	**80%**				
200 TB	40%	Useful power	16,000	40,000	64,000			
		Lost power	24,000	6,000	16,000	**Total Power**	**Annual Cost**	
		SAN 1–5 Power Footprint Totals	**40,000**	**6,000**	**80,000**	**126,000**	**$93,820**	
After Cisco IVR		**Infrastructure:**	Total = Total archive units x Total 400W (redundant) supplies (800W)					
	Utilization	**Tape Archive**	**N + 1 SMPS (watts)**					
SAN 1	70%	10 x Generic	8,000					
SAN 2	70%	10 x Generic	8,000					
SAN 3	70%	11 x Generic	8,000	UPS	CRAC			
		Total power	24,000	36,000	48,000			
Storage Capacity	**Average Utilization**		**80%**	**80%**				
210 tb	70%	Useful power	28,000	24,000	38,400			
		Lost power	12,000	3,600	9,600	**Total Power**	**Annual Cost**	
		SAN 1–3 Power Footprint Totals	**24,000**	**3,600**	**48,000**	**75,600**	**$56,292**	
					Delta	**50,400**	**$37,528**	

The measurement of power per work unit performed is still evolving and requires more development before standard formulas and methodologies can be applied across a data center operation. It is clear that measuring efficiencies in the data center is and will continue to be a top priority for users and vendors. This paper seeks to provide some insight from the perspective of the network, but the in-depth analysis required to establish efficiency of a data center involves a much larger effort. Cisco is committed to working with its customers and partners to work toward the greening of data center operations.

Best Practices and Solutions Summary

This paper provided a consolidated dissemination of various aspects of designing efficiency into a data center operation using Cisco and partner technologies. In addition to the solutions highlighted, many best practices were identified. Table 3 provides a consolidated overview of these best practices, and Table 4 provides an overview of the solutions detailed in this paper. These solutions are being shared as insights and are not intended to be prescriptive.

Table 3. Best Practices

Practice	Description	Effect
Regular utilization audits	• Perform a manual or ideally automated analysis of server, storage, network, and facilities utilization rates. • Identify optimization points as defined by low efficiency rates (low to be defined by operation).	• Data center operations team can set utilization rate benchmarks. • Audits identify optimization points that can open power capacity and set efficiency benchmarks.
Bringing non-IT assets onto the network	• Take advantage of the reach and intelligence of an IP-based network to complement or replace older analog networks.	• This practice offers more visibility into the supporting facilities, enabling better decisions. • This practice allows for monitoring, measurement, and management of the largest power-consuming components.
Organizational transformation	• Develop better working relationships between facilities and IT. • Hire Facilities professionals on the IT payroll.	• Organizational transformation means fewer wasted planning cycles because facilities input is incorporated earlier in the solutions assessment. • The skill base for efficiency planning and analysis is increased.
Governance	• Establish a Change Advisory Board (CAB) to govern efficiency benchmarking, vendor management, skill set dependencies, project funding, and internal and external communications.	• Governance focuses resources on increasing efficiency in data center operations. • Governance identifies and articulates gaps in skilled analysis and planning. • Budget, communications, and reporting can be managed effectively. • Governance provides regular consolidated reviews of new technology and sets purchasing guidelines.
Measurement, monitoring, and management	• Standardize on metrics and management applications. • Implement network-based management applications across IT and non-IT assets. • Integrate network management with server, storage, and facilities management applications.	• Managers can include real data on power consumption, efficiency, and emissions in decisions, trending, and reporting. • Predictive failure and capacity analysis is possible. • Asset management, provisioning, and component availability are improved.
Power profile analysis	• Develop power profiles per component, system, and service across the existing infrastructure. • Benchmark power characteristics and apply standards for new services and infrastructure.	• Power and cooling characteristics are added to new and existing operational profiles, including efficiency and corresponding emissions. • New internal standards are set for procurement, planning, and design of IT services that incorporate power consumption and efficiency.
Green scoring	• Using internally developed standards (today), assign a green score to existing and proposed data center projects.	• An organization's decision makers can consider new, crucial data that can decrease expenses and support strong environmental stewardship.

Table 4. Solutions

Solution	Description	Products
Storage consolidation	Increase the use of storage assets up to 70 percent through the implementation of director-class switching, common fabric, service modules, and IVR.	• Cisco MDS 9500 Series • Catalyst 4948
Application delivery optimization	Provide security, Secure Sockets Layer (SSL) Offload, and load-balancing services through a service module architecture versus one that is appliance-based.	• Application control engine module • Firewall services module • Cisco Catalyst 6500 Series
Advanced services	Cisco and APC/MGE Facilities Assessment professional service assists IT professionals in facilities planning, design, auditing, and general assessment.	• Cisco Advanced Services
Cisco Solutions and Services		
Given the relative complexity of data center operations, each of the suggested solutions introduces power efficiencies into the data center operation. Cisco offers the most appropriate solution to a given customer challenge through specialized partners and integrators. For more information about any of the content in this white paper, please refer to your Cisco partner or account manager.		

For More Information

For more information about Cisco data center solutions, contact your local Cisco account representative or visit: http://www.cisco.com/go/datacenter.

Suggested Reading

- Cisco Press: Build the Best Facility for Your Data Center
- APC Press: System Specification and Project Manual, Volume 1: Optimized for Small and Medium Data Centers
- APC Press: System Specification and Project Manual Volume 2: Optimized for Telcom/Networking/Server Rooms
- Natural Capitalism: Creating the Next Industrial Revolution

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

Asia Pacific Headquarters
Cisco Systems, Inc.
168 Robinson Road
#28-01 Capital Tower
Singapore 068912
www.cisco.com
Tel: +65 6317 7777
Fax: +65 6317 7799

Europe Headquarters
Cisco Systems International BV
Haarlerbergpark
Haarlerbergweg 13-19
1101 CH Amsterdam
The Netherlands
www-europe.cisco.com
Tel: +31 0 800 020 0791
Fax: +31 0 20 357 1100

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Printed in USA C11-418601-00 07/07