ROBERT FRANCES GROUP
Business Advisors to IT Executives
120 Post Road West, Suite 201, Westport, CT 06880 Phone: 203-429-8950 Fax: 203-429-8930

Building a Green Data Center

Jerald Murphy
Senior Vice President and Service Director
Robert Frances Group

120 Post Road West, Suite 201, Westport, CT 06880 Phone: 203-429-8950 Fax: 203-429-8930

Executive Summary

Most IT executives have become painfully aware of the increased problems associated with power and cooling in the enterprise today. While the computer industry has seen the direct benefit of smaller, faster circuits predicted by Moore's Law in terms of price and performance, these benefits are starting to come with a significant cost: power consumption has increased to the point where it costs more to power and cool systems than the money it takes to buy the systems.

This problem has crept into the data center over the last couple years because the direct cost of powering and cooling IT equipment has not been paid by the IT shop. Typically, facilities managers have been responsible for paying this power bill. Up until 3 years ago, electrical usage had been a smaller fraction of overall IT expenditures. Electrical costs were relatively cheap, total consumption was relatively low, and the business focus had been on increasing the flexibility and performance of the data center.

A fundamental shift in the power equation has occurred for several key reasons. First, overall computer performance increased to the point where computation is now the commodity, making it easier to purchase more performance. Second, an increase in the consolidation of servers has increased the density of power needed in the data center. Third, the increasing cost of energy is driving business to find ways to become more efficient to save money. Finally, the global focus on the environment and governmental regulations are putting pressure on businesses to become more environmentally conscientious.

At the same time energy costs are rising, energy utilization has moved from a footnote to a significant operational constraint. In addition to the high cost of electricity, utilities are starting to find themselves in a position where they do not have enough capacity to handle current (and future) data center demands. The challenges are even more severe internationally. In countries like Japan, the cost of energy can be over twice the average rate of the United States. Additionally, government regulations often make it difficult to adjust energy infrastructure. For example, in Germany companies have to get a ‘permit’ from the government for making modifications, and there are no more permits available for several years.

Robert Frances Group (RFG) believes power and cooling optimization will become the most significant IT issue in the data center over the next 3-4 years. Not only is this a major problem for large corporations, but medium size companies are also feeling the "heat" of these issues.

IT executives will need to develop a comprehensive, end-to-end strategy to get ahead of this issue. IT staff and facilities management will need to work together to improve power planning, and IT executives will need to change their overall infrastructure plans to focus on power and cooling optimization. While each area of the data center must be

examined for power efficiency, it is clear that the increasing density of servers in the data center both provide the greatest current limitations, while offering good potential for efficiency gains. IT executives should partner with companies that not only have efficiency built into all aspects of their product portfolios, but also ones that have the experience to integrate facilities and IT issues into building the data center of the future.

Current Data Center Environment

IT executives are finding many data centers are at the limit of their power and cooling capacity. This is a cause of great concern since many of the data centers are only one third to half way through their planned lifecycles. This is not because the data centers have all their racks completely full of computing gear, but rather because there is no ability for the data center environment to provide the energy needed, or to provide the necessary cooling capability to keep existing systems running. Data centers have reached this critical stage through the perfect storm of increased consolidation, exacerbated by higher density electronics and ever increasing energy costs.

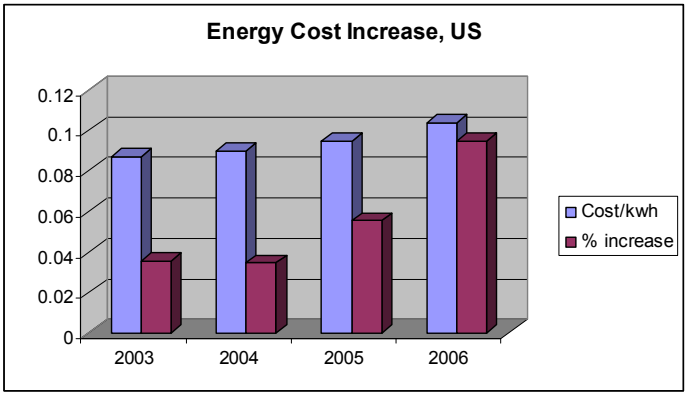

Figure 1 Increase in US Energy Costs Source Energy Information Administration, 6/7/07

Drive for Consolidation

Over the last 10-15 years, IT executives and data center managers have lost control of computing. The era of distributed computing brought the price of computation to the point where any department could use money in its own budget to build its own business applications. Unfortunately, this has brought companies to the point where there are far too many data centers than are needed to run the business. Mergers and acquisitions have situations where the unified company has many applications doing the same function. With web-based computing and an increasing need for application integration, the complexity and cost of these distributed environments has gotten out of control.

Increased Centralized Control

IT executives have been looking to increase the efficiency of IT operations, both to rein in costs and to make it easier to integrate the disparate applications that most companies have today. This increased control will help IT managers better plan future application develop and ensure that different business processes are working together. Most large companies find they have hundreds, and sometimes thousands, of applications. Centralizing the control of these assets makes it easier for IT executives to manage these assets.

Save Money

In addition to taking back control of the IT infrastructure, centralizing IT assets can generally save a significant amount of money. Companies that can cut the number of data centers it is operating can reduce the cost to run the remaining facilities, reduce real estate costs, reduce personnel costs, and lower the utility bills. Another large issue is application and infrastructure rationalization. First, when companies reduce the number of applications running, they eliminate the licensing costs of the application. Second, they reduce the number of personnel needed to support the application. Third, they eliminate the hardware costs and space that these applications need to run on.

RFG recommends that IT managers consolidate many applications onto fewer platforms. Most applications on dedicated servers run at 15 percent CPU utilization and use no more than thirty percent of the storage. Consolidating multiple applications onto one platform can maximize the utilization of a single platform, while further reducing the space and energy requirements of the rationalized systems.

Problem with Consolidation

The good news about Moore's Law is that it has allowed more computing power to be applied for the same price. From 1958 to 1992, inflation increased at an average 3.66 percent, while computer costs decreased on average of 19.13 percent.[1] Unfortunately, chips with faster clock cycles dramatically increase their power consumption, while the heat generated by each chip increases. Even as companies move to dual-core processors with reduced clock cycles, the problem does not go away. While multi-core chips at lower clock speeds may reduce the power per CPU, the total power and heat dissipated by each chip has increased.

[1] "Computers and Growth", Jorgenson, D and Stiroh, K, Econ. Innov. And new Tech. , 1995, Vol3, p 295.

More Systems in Smaller Space

While the CPUs in servers have gotten hotter, they have also gotten denser. This means that more systems are packed in the same space, which means that more cooling is needed. The problem with this is that standard data centers were built assuming a cooling density based on assumptions about air flow that are no longer valid. Most of the data centers that were built as little as five years ago were designed to cool racks of computers with a heat generation of no more the 8-10 kilowatts. Densities associated with current blade systems can generate up to 35 kw of heat per rack.

Power and Cooling Problems

Many company executives RFG has spoken with have said that these new power and cooling problems present several simultaneous problems that are very difficult to solve at the same time. First, the increase of density of these new computing systems cannot be cooled with the existing raised floor air conditioning systems. Second, adding existing air conditioning units to legacy data centers requires space, which is needed for the computers that are occupying the data center as a result of previous data center consolidation efforts. Third, the cooling systems themselves require power to operate, anywhere from 50 percent to 100 percent of the energy required for the data systems themselves[2]. Even if the space could be found for the cooling systems, which it often can't, the utility company is often unable to deliver the power required to run all the computing and cooling, even if the data center manager could afford it, which he often can't.

Existing Systems were not built for energy efficiency

Unfortunately, many of the existing systems in the data center were built for performance, not efficiency. These inefficiencies lay throughout most data center system elements. For starters, most systems have inefficient power supplies. Typical existing systems have power supplies that are seventy percent efficient. This means that thirty percent of the energy delivered to the systems is lost as heat. Using power supplies that are ninety percent efficient will cost slightly more, but can "give back" twenty percent of the power going in to the data center. This inefficiency also drives the cooling costs higher. Next, the CPUs themselves are power hungry. Many of the latest CPUs can consume almost 200 watts of energy. IT executives should look for newer CPUs to have similar processing power, but consume less total energy. Another element that is often overlooked is the hard drives in servers. Hard drives both consume energy, and generate heat in the server. This is especially true for blade servers that have internal drives. Since drives are mechanical systems, they are more prone to failure than solid state systems. Data center managers can greatly increase the reliability of systems by using blade servers without drives, and consolidating storage requirements via storage area networks.

[2] 37% of data center energy goes to cooling systems, 50% energy used by IT systems, "Reducing Energy Costs in the Data Center, Developing a Plan of Attack", BM Magazine, Issue 222, Article 271594

120 Post Road West, Suite 201, Westport, CT 06880 Phone: 203-429-8950 Fax: 203-429-8930

Data Center Doesn't Have Enough Power

Even if data center managers have the space to put in special purpose cooling systems that can achieve the cooling densities required for modern systems, the facilities often don't have the ability to drive the power needed by these denser systems. Typically, the facility inbound power distribution is designed by facilities management. The rooms that have the power distribution units (PDUs) were built to handle an expected power load. When the power densities of modern blade servers are brought into consideration, the data center PDUs are often maxed out, with new space needed to condition additional power brought from the utility company. Not only is more space needed for additional PDUs, but additional space is required for the uninterruptible power supplies (UPS) that serve as backup power. The space is not usually available for the required augmentation, and RFG has found that utilities frequently don't have the power available to deliver to the data center, even if the above allowances have been planned.

The power required is not just for the IT systems, but for the cooling units as well. Most companies plan for a 50 percent increase in the power needed to run the cooling systems. However, RFG has found that the power needed by cooling systems can equal that required by data systems. In addition to the power needed by cooling systems, the utilization of cooling is just as problematic as the utilization of power. First, the airflow itself needs to be managed. Raised floors in data centers have been designed to deliver cool air. However, a good portion of the area under the raised floor is often blocked by cables and other connectors, limiting the effective air flow. Additionally, air is lost through leakage in the floor, and systems at the top of racks often don't get the same cold air that is delivered near the floor. These problems are compounded by the historic lack of communication between data center and facilities management. Facilities managers are not usually in the same organization as data center managers, and liaison between the offices has not been well coordinated.

Most data centers are not able to be fully powered, which prevents them from being fully utilized. Compounding this is the fact that data centers are not effectively providing the right amount of cooling to the right areas. This then requires cooling to be over-provisioned, in order to meet all data center cooling requirements. The waste associated with the inefficient utilization of energy not only costs companies more money, but negatively impacts the enterprise's environmental sensitivity.

120 Post Road West, Suite 201, Westport, CT 06880 Phone: 203-429-8950 Fax: 203-429-8930

Requirements for the Data Center of the Future

In order to create a positive change in future data centers, IT executives need to take the initiative to work with facilities managers and IT vendors to save money and increase efficiency. The data center strategy should be looked at holistically, not simply from an IT standpoint, but from the overall facility management perspective as well as IT's data center requirements.

Environmentally sensitive planning

One of the first things data center managers can do is to start with environmental planning. IT executives should work with facilities planners as well as legal and real estate planning to determine the best location for future data centers. Not only should an area be considered where the unit cost of energy is low, but should also be an area that works for data center survivability (i.e. out of hurricane zones, earthquakes, etc.), but will also be a central location from a business process perspective. This will require IT executives work with line of business managers to understand application process flow. In addition, the coming importance of global warming has increased the desire of some companies to bring use eco-friendly renewable power in their data centers. Not only is renewable power environmentally conscious, but its use can ensure a higher degree of overall energy security. For example, RFG is seeing an emerging trend of data centers moving to locate near hydroelectric power. IT executives should work with companies such as HP that have the understanding of these complex factors to help give advice on how to integrate these issues to arrive at an optimal data center solution.

Once optimal locations have been determined, thought should be given to power distribution. Coordination should be made with utility companies to more accurately predict future power needs, and facilities management should have scalable plans for future PDU requirements. Some consideration should be given to where and how AC/DC power conversion should be done. One approach to gain higher efficiency would be to run 480V, 3-phase AC to the rack. Additional design thought should be given to the tiers of power. Many data centers are adding another tier to UPSs with flywheels, that can eliminate the very small interruption cycle times that eat away at battery backup life.

These issues can only be effectively planned for in the budget, design, and space with tightly integrated planning and coordination between facilities and data center management. RFG is starting to see some companies that have a facilities planner as a member of the data center staff. Whether or not there are organizational changes, it is clear that close communication exist between these groups. The design engineers that work on these issues must understand the environmental, power and cooling issues associated with all aspects of the data center.

Energy Efficient Environment

Data centers of the future must be much more environmentally friendly. The servers that are used should be designed in ways that are more efficient. HP is one of those vendors who are designing energy efficient servers. HP leverages low power memory, low power processors, and low power small form factor drives to achieve optimal energy efficiency. With most servers operating below 15 percent utilization, an emphasis should be placed on CPUs that use the least amount of power. Power supplies should be high efficiency, and eliminated altogether if possible. The power that is used by all systems must be managed more efficiently.

IT executives should use appliances that can monitor the environmental conditions, like the HP iLO2 and power manager software. And, control systems should be employed that can dynamically change the power and cooling envelope to meet the actual business computing requirements. Controls should ideally be placed at every point in the data center environment, from controlling the temperature of the room, the air flow of cooling units, to the speed and voltage of the computing systems. Hard disks can be designed where the default is that the disk is not spinning, and only turn on disks when absolutely necessary.

In order to achieve the consolidation necessary for increasing system and storage utilization, virtualization technology will need to be used that can dynamically adapt the computing and storage workload to changing business conditions. Ideally, these adjustments are made automatically.

Power generation should be more efficient, so less energy is lost to power conversion. Additionally, power systems should be designed for redundancy, at every level. That is, redundancy for power coming into the facility, redundant power distribution, and even redundant system power supplies. The more heat that can be taken off of system boards, the less heat that needs to be dissipated and the higher the system reliability.

RFG believes data center managers should need to optimize space utilization. Many racks are "maxed" out when only fifty percent full and planning time should be spent on increasing space efficiency throughout the data center.
RFG recommends consolidation storage to a storage area network. This will allow the consolidation of data, which will both reduce money spent on storage, and increase available data center space. One way that storage can further be optimized is through the use of "thin provisioning". This is a technique that essentially "over provisions" the use of storage by multiple applications. Then, disk space is only "given" to an application when it is actually required on a write. Tests have shown that this technique can improve storage utilization from four to one hundred fold. However, close attention needs to be paid to storage systems, so that critical applications never have a chance of not having storage space when it is needed. Another new technology that can improve storage efficiency is the massive arrays of inactive drives (MAID). This technique employs large

data caches, which allow storage systems to be in the default "off" mode. This dramatically decreases the energy needed by storage systems, at the cost of a small increase in data seek times if the data is not available in the cache. IT executives should evaluate new technology based on these new energy-efficient metrics.

Additionally, networking components should be integrated into the overall systems architecture. One of the benefits of bladed environments is that they have networking infrastructure integrated into the blade chassis. This saves space, reduces the money that needs to be spent on networking equipment, and is designed to be tightly integrated into the systems architecture.

Finally, next generation data centers need to have a well-designed management environment. It is not enough to know that systems are up or down, or even to automate the provisioning of resources. Data center resources of the future need to be dynamically adaptable to changing power requirements. Since energy costs are only going to increase, it is important that management systems understand the current energy environment, while providing adaptive feedback mechanisms to change the power profile of systems to dynamically adjust to changing enterprise demands. IT vendors are starting to develop systems that both have energy information embedded for management, but are also providing interfaces to changes power at the chip level. Future systems will need to be designed so that voltage and clock speeds can be adjusted to use a little power as necessary to meet with business requirements. In multi-core processors, this will mean eventually giving the capability of turning off processors that are not needed. IT executives should look at all aspects of design, implementation, and operation of data centers to find ways to decrease energy and cooling costs, which will usually result in more effective processing, as well as improving the corporation's "green" image.

Call to Action

While it's great to have plans for an integrated, efficient, environmentally sensitive data center of the future, almost all companies already have data centers that have systems that cannot simply be thrown away. Because of this, IT executives should look at what they can do with what they have today? RFG believes there are many actions that can be taken immediately to improve the efficiency of existing data centers. IT executives should evaluate each of these suggestions and prioritize which actions will yield the most benefit in the shortest time.

Application Portfolio management

One of the most important actions that IT executives can take is to evaluate the existing set of applications to determine whether any applications can be retired. Often, data center consolidations followed by application audits will find many applications that are no longer needed, either because they are not used any more, or are a duplication of another application running. When companies get rid of applications that are no longer needed, the machines these applications are running on can be turned off (assuming it is an application on a dedicated system, which most are). This saves 100 percent of the

energy required to run the system, another 50 percent of that energy that was used to cool the system, as well as the space that system occupied. Not only does this action save the energy costs, but can often eliminate application licensing and maintenance costs.

Improve use of existing infrastructure

Most of the infrastructure in the data center can be used more efficiently, if a thorough audit is done of the environment from an energy perspective. Consolidating underutilized servers will improve efficiency and save space. This usually requires adding some virtualization software to allow the sharing of resources. Additionally, the virtual servers that exist will still require administrative support, plus the additional support required to administer the virtual environment. RFG finds that companies that can achieve a 5:1 server consolidation through virtualization will usually save money.

Another action that can be taken immediately is fixing poor environmental circulation factors, aligning them with data center facilities best practices. Devices can help separate hot air from cold air, including blanking panels, insulation kits between bayed racks and insulation panels installed at the top of the rack. Blanking plates can be purchased and put in any spaces in the racks to optimize air circulation. Raised floor areas can be cleaned to eliminate any wasted cabling and any air "dams" that lower the cooling efficiency. "Hot" resources should be optimally placed to make sure the "dense" systems are optimally cooled. Data center rows should be designed so that every other aisle is hot and cold, to allow any augmenting cooling units to concentrate hot air evacuation. Some companies, such as HP, have capabilities such as Dynamic Smart Cooling (DSC) that can be easily added to the data center to track temperatures and automatically adjust the fans and cooling units to alter air flow based on changing data center requirements. This type of capability can reduce cooling costs from 25 to 40 percent.

Finally, IT executives should have team in place that understands how to pull all the pieces together, while working with what resources currently exist. There should be someone from the IT staff that has a primary responsibility to focus on these issues. In addition, data center managers should call on the assistance of outside consultants who have the experience dealing with these energy issues in a wide variety of conditions to help bring these benefits to today's data center as quickly as possible.

HP has products and services that are focused on energy efficiency and environmental consciousness. They have experienced staff that can help companies with their short-term tactical fixes, as well as long-range strategic planning and implementation. HP data center experts can analyze facilities to find where the areas where efficiency can be improved. Products, like Dynamic Smart Cooling, can be applied to make more efficient use of existing resources. HP has a Modular Cooling System that can increase the thermal load handled by systems up to 35kw. Finally, HP power efficient IT systems can be implemented to further increase data center system efficiency.

Conclusion

Data center managers are under a lot of heat to reduce IT operational costs, while at the same time improving infrastructure performance and being environmentally friendly. Ignoring this issue will not be an option. While there are many benefits to reducing the size and increasing the speed of computers, reducing power consumption has not been one of these benefits.

IT executives need a new approach to designing the data center of the future. First, they need to work in close cooperation with facilities managers to do more effective long range planning. Planning should be done for the layout of the facilities, the efficient distribution of power, and the effective elimination of heat. At the same time, IT designs need to be reconsidered that are holistically integrated, with efficient resource utilization, with the dynamic ability to redirect computation, power, and cooling as needed to maximize performance at the minimum cost.

There are immediate actions that can be taken today to lower energy costs and gain back valuable data center space. However to take these actions, data center managers are going to need the experience of people who have seen these challenges before and can tie together the tactical improvements with a long-term efficient data center strategy. HP is one company that has the experience to help companies throughout the entire spectrum of data center infrastructure. HP systems are built with energy smart designs. HP's world class management systems can monitor the infrastructure and change systems on the fly to adapt to changing situations. HP has experience with both data center and facilities systems, in order to build and implement a green environment from "chip to chiller". IT executives should place a high priority on improving the efficiency of their data centers, and look for partners that have the products and practices to help achieve these ends.