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

Breaking the Power Deadlock:

The Power and Cooling Benefits of Running on Open Systems

White Paper

Published: July 2008

For the latest information, please see www.microsoft.com/mainframe

© Robert Frances Group, 2008

Contents

Executive Overview

For the last several years, electrical power issues in the enterprise have been rapidly escalating. Although industries have seen the direct benefit of Moore's Law with respect to the price/performance curve, that same law possesses a significant downside that companies have not noticed until just this past year: until recently, power consumption has increased at the same exponential rate that performance has increased.

IT executives had been largely unaware of this phenomenon until recently, because the direct cost of this problem is not borne by the IT shop. Typically, facilities managers are responsible for paying the electric bill. About three years ago, electrical usage was considered a commodity. Electrical costs have been relatively cheap, total consumption was relatively low, and the emphasis in the business has been primarily on increasing performance.

The past year has seen a fundamental shift in the performance metrics for IT in the enterprise. Performance of integrated circuits has improved to the point that raw computation is now the commodity. Across many industries, the average utilization of most processors in the data center is 15 to 20 percent. At 20 percent utilization, there is not usually a compelling need for the fastest processor, but rather a need for the processor that can efficiently meet business objectives.

At the same time, electrical power has moved from being a commoditized utility to being an operational constraint. The cost of electricity has started to skyrocket, along with the rest of the energy industry. Utilities often do not have enough capacity to satiate the ever-increasing demands of the data center. Finally, the typical data center does not usually have the physical capability to handle the power and cooling requirements of a building full of blade servers, even in cases where the company can afford the cost and the utility industry can provide the power.

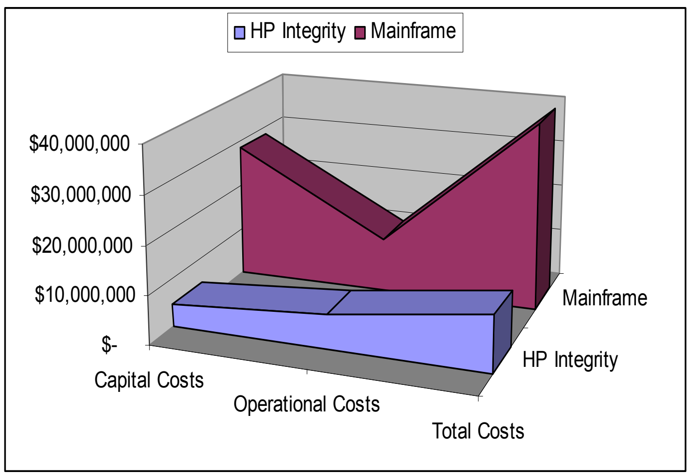

Figure 1. Cost Comparison of IBM System z to Hewlett-Packard Integrity (Source: RFG)

Robert Frances Group (RFG), a leading research and consulting firm focused on the IT workplace, believes power efficiency will become the number-one IT issue in the data center for the next three to four years. RFG has interviewed many IT executives since 2005, and has found power already to be the primary IT issue for many large financial institutions.

To get out from under this enormous difficulty, IT executives will have to tackle it from every angle. Not only do IT and facilities management need to work more closely together for power planning, but IT executives must shift their computing criteria priorities from price performance to power performance. Each area of the data center must be examined for power efficiency, but it is clear that the increasing density of servers in the data center provides both the greatest current limitations and the greatest potential for efficiency gains. To get the greatest return on investment (ROI), IT executives should look closely at those server manufacturers that focus on power efficiency.

© Robert Frances Group, 2008

IT executives should look not just at cost reductions for these systems. They should also understand the ROI domino effects that grow out of such power efficiency moves. Among the other benefits within the enterprise are reductions in data center space, reduced data center and facilities management costs, lowered system failure rate, and more efficient space utilization. In the case study examined by RFG, a Microsoft®/HP technology combination demonstrated significant savings. The capital savings for using HP Integrity servers was over U.S.$23 million, while the operational cost savings was well over $4 million.

Overview of the Current Data Center Environment

Inefficiencies of Current Servers

Servers that currently reside in data centers are not as energy-efficient as they should be. This is a result of companies focusing primarily on price and performance for the last ten years. This emphasis is certainly natural, as the total cost of power for running and cooling hardware over the previous ten years was only a small fraction of overall operational costs.

However, in the last two to three years, the situation has changed dramatically. The density of computing systems continues to be a marvel of technology, bringing the cost per transaction continually lower, from a cost of computing standpoint. Unfortunately, the amount of heat generated on ever-smaller devices has geometrically increased the amount of money spent on providing power to the data center, including money spent on cooling the hardware in the data center.

This differential has gotten to the point where it can cost more to run and cool devices than to buy the devices in the data center! Following is an overview of each element in the data center that is exacerbating this problem.

Servers

The server is the key element in the data center—it does the work of business. It is the one item in the data center that consumes the most power when running and requires the most power to cool as well. Although many people consider the central processing unit (CPU) as the main element that consumes power and generates heat, several other elements contribute to overall inefficiency.

Power Supplies

The power supplies used in most servers today are relatively inefficient. Typical server power supplies for standalone servers are only 70-percent to 80-percent efficient. Given that a large data center can spend over $1 million per year on energy costs, the use of inefficient power supplies can represent up to $300,000 per year in wasted energy. Robert Frances Group (RFG) believes that IT executives should insist on using power supplies that are at least 90 percent efficient, for all data center servers. The higher-efficiency power supplies will cost more money, but companies will more than make up for the price difference in reduced energy costs.

A typical server uses an average of 205 watts (W). For 24-hour-a-day operation, the daily consumption is 4.92 kilowatt-hours (kWh)[1]. At 11 cents per kWh, the annual energy cost per server is $197.54. A typical high-efficiency server power supply will cost an additional $30[2], but with an increase of efficiency of 25 percent, will save $49.38, with an ROI of just over 7 months.

---

[1] Page 5 of Green Grid White Paper "Five Ways to Save Power" www.thegreengrid.org/gg_content/White_Paper_7_-_Five_Ways_to_Save_Power.pdf

[2] "Google, Intel target efficient PC power supplies" http://news.cnet.com/Google,-Intel-target-efficient-PC-power-supplies/2100-11392_3-6190576.html?hhTest

© Robert Frances Group, 2008

Chassis

Server chassis can be a source of wasted energy and inefficient heat management. Many server chassis in use today were designed for cost savings, not energy savings. Inexpensive, unreliable fans, inefficient heat sinks, and poor designs for total air flow are typical attributes that can be found on legacy servers.

In addition, server chassis that have not been designed for redundancy, coupled with poor cable management, together serve to restrict airflow, reduce server life, and potentially put mission-critical systems at risk for failure. IT executives should look for modern server chassis to have easily replaceable components, with high-efficiency fans, effective cable management, and intelligent designs for airflow to increase system availability and efficiency.

Heat and Cooling Mechanisms

Many servers have not been properly designed for moving heat away. Design flaws include blocked airflow, as well as no way to change the attributes of a server to reduce energy consumption based on changing workloads. Older systems have no way of increasing the rate at which heat can be evacuated from a system. With current blade servers, which can achieve power densities of up to 35 kilowatts (kW) per rack, it is essential to have advanced mechanisms in place for cooling servers, because standard raised-floor rack environments can cool only up to 11 kW per rack at best. RFG suggests that companies use server systems that can communicate with rear-door heat exchange units, a strategy that can augment raised-floor cooling to achieve higher cooling densities.

Storage Resources

Just as most servers in use at today's data centers operate at low CPU efficiencies (typically less than 20 percent), those same servers are also very inefficient at using their available storage resources. Dedicated disk utilization of servers is typically below 20 percent. Maintaining storage directly on the server using a direct access storage device (DASD) leads to increased heat (at least 10 watts) and difficulties in maximizing cooling efficiency for the server (due to the physical space taken by storage that blocks or limits air flow. Reliance on DASDs also decreases the life of the server, because as mechanical devices, they are among the first elements to fail in a server system. RFG recommends that IT executives move to using storage area networks (SANs) to more efficiently use storage resources.

Impact of Electrical Power Issues

As a result of the inefficiency of current server systems, the data center is currently in crisis. Any one of the issues discussed earlier can be significant. However, given the increase in power density at the data center, as well as the troika of space limitations, cooling inefficiency, and insufficient management coordination in data centers, servers are bringing hard limits to data center capabilities and geometric increases in data center costs.

Space limitations. Even though many data centers have the physical space for more servers, that extra space often cannot be used because the facilities either cannot provide the energy to the space or the existing cooling systems cannot cool to the density required. If a rack of blade servers needs 35 kw of cooling per rack, but the air conditioning can provide only 11 kw per rack, then these racks can be filled to only one-third capacity, leaving two-thirds of previously planned space unavailable for computing. RFG has talked with many company data center executives who found they needed to invest millions of dollars years ahead of schedule because space they had planned to fill for ten to fifteen years was effectively unusable after three to five years of service.

More frequent failures. Sometimes, companies try to use more of their available space when the cooling metrics do not support this decision. When this happens, the results are almost always unfavorable. The mean time between failures (MTBF) of IT systems drops precipitously when racks are filled beyond their cooling capacity.

In some cases, the MTBF decreases even when densities are within acceptable ranges. This decrease is often because of poor rack cooling design, or because there are gaps between the servers in the rack. Such an arrangement creates local cooling "loops" that trap cold air in the lower half of the rack, preventing systems in the top half of the rack from being properly cooled.

Insufficient management coordination. One of the biggest issues concerning the limitations of current data center environments is lack of coordination between facilities managers and data center managers. Facilities managers are usually responsible for the physical space of the data center, including the design and placement of air conditioning throughout the facility, the energy coming into the building, and coordination with the local power utility.

The data center manager is responsible for all the IT systems in the data center. These systems generate the requirement for power that typically represents more than 80 percent of the energy used in many data centers. When these two managers are not in close coordination, data centers are susceptible to insufficient cooling capacity to meet data center requirements—or even to running out of power. When data center facilities are built in life cycles of fifteen to twenty years, yet power requirements double every eighteen months, it is easy to see how long-term planning, without detailed and continual communication between these two individuals, can lead to disastrous results—and usually does.

Initially these issues were felt only by the largest companies with the largest data centers. RFG has found that power and cooling issues have progressed to the point that even small and midsize businesses are feeling the effects of these issues in the data center.

RFG recently talked with a small manufacturer of ink supplies that had only three racks of equipment in its data center. Even so, the density of the manufacturer’s machines had reached the point where critical business systems were shutting down at key times during the year. The dramatic increase in system failure rates was costing the company thousands of dollars in lost business, and hundreds of hours of lost productivity as IT staff struggled to keep these systems up.

RFG finds that even in smaller companies, server consolidation can still be beneficial, in terms of both reduced operational costs and efficient powering and cooling of all the systems in the data center.

Companies of all sizes need to embrace next-generation requirements for operating environments, RFG believes, to reduce operational costs, increase business flexibility, and improve the reliability of all servers. The initial power cost of any given system is usually less than the software costs associated with that system, but energy costs will continue to rise. The cost of oil as of 22 May 2008 was at an all-time high of $133.17[3], and it shows no signs of decreasing. This constraint is making managing total power throughout the data center ever more critical.

Requirements for Successful Operating Environments in the Enterprise

In order to have successful operating environments, RFG believes that IT executives need to establish data center standards that improve energy efficiency and take advantage of virtualization and scalability, while creating an environment that is both manageable and easy to use. Requirements for these areas, as well as integration capabilities and decreasing the total cost of ownership (TCO), are discussed in this section.

[3] *Wall Street Journal,* 22 May 2008.

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

Energy Efficiency

The most important new requirement that IT managers must look at is energy efficiency. IT executives should add energy efficiency requirements to all future systems service contracts, according to RFG. RFG predicts that it will be only a matter of time before governmental agencies start legislating data center efficiency. In addition, increased energy efficiency will dramatically lower almost every company's operational costs, so it also makes business sense to pursue these goals. Here are four suggestions for helping to decrease those costs:

More efficient power supplies. Insist on purchasing power supplies that are at least 90-percent efficient. The less power that is lost through power conversion, the more money saved. Having fewer power supplies that are more efficient and available will save money as well. This is best done with blade servers, because multiple blades can share central power supplies.

Newer-model chips. Purchase computer chips that are designed for lower power consumption (less than 100 watts per CPU). In addition, and perhaps more important, choose chips that can dynamically change the power consumption based on the workload presented. In close coordination with your chip selection, check that the software vendor offers an operating system that can interact with the chip's power management features, and management software that can monitor overall power conditions in the environment (such as average power, peak power, and temperature) and make changes as needed to optimize power utilization.

Improved heat dissipation. Ensure that your next-generation servers can dissipate heat efficiently. You should be able to modify cooling based on load, in addition to interacting with higher-density cooling capabilities such as rear-door air conditioning, or in-row air conditioning units. The cooling components themselves should be optimally efficient, with efficient fans that consume less energy than their predecessors, are more reliable, and remove more heat from the servers.

Refined system design. Build the server for efficiency from the ground up. Include planning for integration with virtualization platforms, having server components that can be hot-swapped, close integration with management software, and incorporation of corporate recycling programs.

Virtualization Capability

Next-generation servers should make the most of virtualization's advantages, while taking into consideration the potential challenges that are associated with virtualization (including management, complexity, and security). Some of the benefits of virtualization include:

Allows easy consolidation. Server virtualization can make it easier to consolidate applications. RFG believes that virtualization is simplified when architectures use standards for common application environments. Along with appropriate tools, virtualization can be easy to use and manage. RFG believes this happens best when using a common operation environment.

Simplifies scalability. Add new workloads to the business environment without having to purchase extra hardware for every new workload that comes along. Instead, gradually add new processing power, RAM, and storage in a measured, predictable way, independent of any one application.

Improves integration. Integrate applications into external application environments. As companies begin to develop service-oriented architectures, this capability is becoming essential. When applications are bound into legacy environments, such as mainframe systems, interconnecting new applications and partner applications becomes more difficult. Operating environments can now work with leading virtualization platforms and facilitate application interconnection.

Permits transparency. Properly implemented virtualization is transparent to anyone who uses the virtualized applications. That is, ideally the users of applications that are on virtualized platforms

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

should be able to use these applications as if they were running on dedicated servers, without rewriting applications.

Scalability

Next-generation IT systems should have high scalability. These systems should be able to handle ever-increasing workloads. In addition, they should permit incremental addition of infrastructure in a way that is transparent to the end user. Such increases should minimize any increase in software licensing, so that more work can be accomplished at a lower cost per transaction.

Ease of Use

One of the key requirements for next-generation servers is ease of use. As systems increase in density, especially with the use of virtualization, they must be easy to put into production and for use on a daily basis. End users should be able to easily take advantage of the new systems. Consider incorporating familiar operating environments with standard interfaces that can remain the same as systems are upgraded.

Manageability

New systems should be easy to manage as well. Manageability extends throughout the entire product life cycle, which includes provisioning service, configuration management, fault management, performance management, and security. Look for systems that can tie into existing management frameworks, and have configuration and performance management that is easy to configure and use. RFG finds that advanced software and processes for managing IT servers that require detailed knowledge and setup often go unused. Therefore, consider server systems that use standard management techniques and can be used in common operating environments. Also, look at implementing systems that can reduce the number of people required to manage the data center environment.

Ability to Integrate with Other Systems

As companies develop more applications with more partners faster, it becomes essential to have systems that are able to integrate both vertically across the enterprise IT infrastructure, as well as across systems of other partners. For this purpose, consider well-established platforms that are used by the majority of companies, as well as software that is well known and protocols that are widely supported.

Low TCO

When looking at the cost of running next-generation data centers, do not look at the purchase price alone, but evaluate the total cost of ownership (TCO): the price of running systems across their entire life cycle. This includes the cost to purchase hardware and software (capital costs), and operational costs as well, such as the maintenance costs for hardware and software, as well as the personnel costs associated with running the systems and the applications that are running on them.

Finally, pay close attention to the energy costs related to running these systems, as well as the costs associated with the space that the hardware occupies. The impact of these costs will be influenced by the geographic region where systems are located. However, with the rapidly rising price of energy, the power costs will only get more significant for *all* regions and companies over time.

© Robert Frances Group, 2008

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

Sample Application in a Next-Generation Data Center

For a full assessment of all the costs for a given application use, IT executives should consider both capital and operational costs of a mainframe environment, as compared with an open systems environment. To put these next-generation data center costs into perspective, RFG looked at a typical use case. In reviewing a case study, the various costs that have been described in this paper can be evaluated in a more concrete manner[4].

Cost Matrix Defined

Costs consist of both the costs of acquiring the solution—the capital costs—and the costs of running the solution on a daily basis—the operational costs. These broad categories are further broken down as follows:

Capital costs (one time)

- Hardware –Computer equipment, networking equipment, and cabling
- Software –Acquisition of operating system software and application software
- Facilities – Physical data center space, racks, power distribution, air conditioning

Operational costs (annual)

- Maintenance – Maintenance fees for the hardware and software
- Personnel – Salaries of employees who maintain the computer systems
- Electricity – Power to run the computer systems and the cooling equipment

Within the operational cost category, the cost of electricity is significant, because in some cases the expense of running the cooling equipment can be almost as high as the expense of running the computers.

Moving from the Mainframe

In this example, a client migrated an SAP system from the legacy mainframe environment onto a platform that consisted of HP Integrity servers with a Windows Server® operating environment. The data in Tables 1 and 2 comes from actual numbers used by the client.

The SAP system was moved off a legacy IBM system z 2094 Model 720 mainframe[5]. The applications were migrated onto two HP Integrity rx8640 servers and two rx7640 servers, all of which were running Windows Server 2003.

The rx8640 has several features that allow for high availability and ease of use. It contains a bank of hot swap fans in the rear of the unit that allow for easy replacement, as well as redundant hot swap power in the bottom rear of the unit. The typical power consumption for these units is 1870 watts. The power consumption for the IBM 2094 Model 720 mainframe is 18.3 kilowatts.[6] The space used by this system is 2.5m^2.

---

[4] Information collected from Alinean, Inc. Mainframe Migration study http://www.alinean.com/PDFs/Intel-Mainframe_Migration-TCOStudy.pdf

[5] According to www.tech-news.com/publib/pl2094.html. This system uses approximately 8244 MIPS (1411 MSUs).

[6] IBM mainframe reference chart, www.tech-news.com/publib/pl2094.html

© Robert Frances Group, 2008

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

Table 1. Cost of SAP on IBM Mainframe

Capital Costs	Dollar Amounts (U.S.$)
Hardware	$3,340,000[7]
Software	$24,800,000
Facilities	$177,043
Operational Costs	
Maintenance	$10,000,000
Personnel	$1,120,000
Electricity[8]	$70,536

Table 2. Cost of SAP on HP Integrity Servers
with Windows Server 2003

Capital Costs	Dollar Amounts (U.S.$)
Hardware	$2,574,000
Software	$1,962,000
Facilities	$95,740
Operational Costs	
Maintenance (equipment and personnel)	$5,244,000
Migration costs to move off the mainframe	$1,600,000
Electricity[9]	$28,831

Comparison of Mainframe to Open System

In this customer's case, the software costs become dramatically lower on the HP servers, in part because the mainframe licensing fees can be very expensive. Looking more closely at the facility costs, specifically energy and space, RFG noted that the HP hardware consumed 59 percent less energy and used 48 percent less space, for a total savings of $123,007 over four years.

In addition, the hardware savings in such a migration are impressive. In this example, the customer had a negotiated mainframe hardware cost of $3,340,000. However, the actual list price for this system is closer to $12 million. Although the capital cost savings for hardware was 23 percent, from list price the capital savings are calculated at over 78 percent.

[7] This is the price paid on customer negotiation. System list price for configuration is approximately $12 million.
[8] Energy costs are computed based on: 18.3 kw per system, one system, running continuously for four years at $0.11/kwH.
[9] Energy costs are computed based on: 7480 watts per system, four systems, 24x7x365 for four years at $0.11/kwH.

© Robert Frances Group, 2008

There are situations in which it makes economic sense to use a mainframe. When there are multiple applications that combine to drive a high-CPU utilization, then the cost of a mainframe might be justified. There are also situations for which there are legacy custom-built applications that would cost a great deal of money to migrate onto open systems platforms. In both of these cases, it is likely going to cost less money in the long run to maintain these applications within a mainframe environment. However, with the new servers like the HP Integrity family, it is possible and increasingly common to carve up the servers via virtualization and hardware partitioning to drive improved utilization on Intel-based hardware. The original design goal for mainframes, over thirty years ago, was to support multiple workloads. However, with resource managers, partitioning, and virtualization, this is no longer the exclusive domain of the mainframe.

Further, RFG sees many companies doing long-term planning that involves retiring legacy applications, and either buying packaged applications, such as SAP, or developing new applications on open frameworks such as Java and Microsoft .NET. Most of these applications show more flexibility and more economic scalability when implemented on open systems frameworks such as HP Integrity servers and Windows Server operating systems. This hardware-software combination has been shown to dramatically lower the TCO in situations that companies frequently find themselves in, such as implementing prepackaged applications that are designed to progressively scale in adoption to changing business needs.

When this is done on HP Integrity servers in a Windows Server operating system environment, studies have shown that the overall TCO can be significantly lower than in the mainframe environment, especially when power and cooling costs are taken into consideration. HP's ability to work with Microsoft in presenting energy information, especially with the new service management solutions from Microsoft, can help lower management costs for the data center.

Finally, given the significant expected increase in energy costs, the benefits of energy-efficient systems will multiply with rising energy costs. In the example presented here, if the energy costs rose to $0.27 per kwH (which is already the actual cost in some locations, such as Japan), the total savings would rise from $41,705 to $102,366.

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

Conclusion

Electrical power metrics are the new arbiters of value for next-generation data centers. Although more advances continue to be made in computer processing power, the increased cost of energy, plus the limitations and costs related to space, work together to make energy efficiency an essential business factor that data center managers cannot afford to ignore. With this in mind, IT executives must begin to look at power costs in their TCO calculations. Most companies are finding that these costs are taking a bigger "slice" of the operational budget. IT executives need to understand how these costs will change over time, and how the concomitant increase in these costs will affect current and future data center operations.

A holistic view of the entire data center operation is required, to optimize TCO. Every situation is going to be different. IT executives should understand business requirements, application architectures, and application integration requirements, in addition to the legacy environment, to determine the best combination of hardware and operating system for reducing their operations costs today, while at the same time flexibly adapting to future business demands.

In some cases, maintaining legacy systems on mainframe equipment makes the most sense. However, RFG finds that there are many times when the combination of open systems, based on HP Integrity servers and Microsoft software, delivers the optimal combination of low acquisition costs, low administrative costs, and an efficient use of electrical power that make a compelling TCO argument.

IT executives should not make simple assumptions based on any high-level marketing information, but should compare their existing operational environments with careful estimates of future demands, to determine the best mix of hardware and software that will allow them to dynamically adapt to current and future business needs. RFG finds that many future environments, especially those that involve prepackaged applications, growing business demands, and flexible integration requirements, yield the lowest TCO when implemented on a combination of HP hardware and Microsoft software. As this paper shows, this can be the case even when adding in the migration costs associated with moving applications off legacy mainframe applications.

Many analysts and vendors will try to convince IT managers that mainframes will always offer the lowest TCO, or that distributed systems will do the same. The reality is that each system can have its value, but the particular environment in which the system is running needs to be evaluated to understand which type of system will provide the best TCO for a given situation. In many cases, the HP Integrity servers running Microsoft software offer this compelling TCO. This value can especially be realized when migrating packaged applications onto lower-cost, highly efficient next-generation systems such as those provided by Hewlett-Packard and Microsoft.

© Robert Frances Group, 2008

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

All rights reserved. 120 Post Road West, Suite 201, Westport, CT 06880. Telephone 203/429-8950. Facsimile 203/429-8930 www.rfgonline.com. This publication and all publications may not be reproduced in any form or by any electronic or mechanical means without prior written permission. The information and materials presented herein represent to the best of our knowledge true and accurate information as of date of publication. It nevertheless is being provided on an "as is" basis. Reprints are available for a nominal fee.

© 2008 Microsoft Corporation. All rights reserved.

Microsoft and Windows Server are trademarks of the Microsoft group of companies.

All other trademarks are property of their respective owners.