

The Dollars and Sense of Virtual Data Centers

By James Geis and Steven Harris, Forsythe

To manage server sprawl, exponential data growth, power consumption, and nearly unmanageable infrastructures, many IT organizations are turning to data-center virtualization and blade-server technologies. But careful consideration is needed.

It's a classic dilemma: The more automation a business requires, the more complex and unwieldy its IT infrastructure becomes. And things seem particularly acute now when business operations hinge on a myriad of interdependent software-driven services, each tied to a unique revenue-generating purpose. These applications, in turn, depend on massive computing resources, not to mention service-oriented architecture and IT Infrastructure Library processes to ensure an extensible, agile platform.

The result? Server sprawl, exponential data growth, power consumption, and nearly unmanageable infrastructures—in short, a data-center environment spinning out of control. To manage it all, many IT organizations are turning to virtualization and blade-server technologies.

In many ways, these innovations are ideally designed to address server proliferation. Virtualization consolidates processing power into a compact physical footprint, frees up floor space, reduces network connections, and brings the administrative and operational efficiencies of fewer racks, wires, cables, and power supplies.

In addition, the technology holds out the promise of a stable, platform-independent approach to legacy-application hosting and better tools to monitor and calculate chargebacks. Virtualization also lets central data centers provide consistent and adaptable service levels to application and business units that want control of their servers—that is, their operating systems and applications—without having to support separate, disparate platforms.

Server virtualization gives IT administrators a view into the underlying physical server resource pool, plus an unmodified OS that allows computing-resource levels to be increased or decreased on demand to meet variable processing-capacity requirements such as higher end-of-month or end-of-quarter needs. While you still have to support multiple operating systems and applications, there

are fewer physical resources to manage and the hardware platform is more fully utilized.

Toss in blade servers from Hewlett-Packard, IBM, Sun Microsystems, and others, and you have the ability to compact even more virtual machines, or separate OS instances, into a platform that provides flexibility and intrinsic healing capabilities. These blade platforms are expected to offer a diskless option by 2008, with SAN and network switches integrated, and will use more networked storage arrays, providing an opportunity to combine and more efficiently handle backup, replication, and disaster-recover requirements through the storage infrastructure.

If it all sounds too good to be true, it may be. Blade servers have their downside: They consume more power and generate more heat than most enterprise data centers were ever meant to handle. The average facility was built nearly two decades ago, when water-cooled mainframes were the rage, not air-cooled microprocessors and disks. If your data center hasn't begun to keep pace with new technology, blade servers may be too much, too soon.

We've studied the pros and cons based on experience with our clients to see whether blade servers are solving or creating problems in the data center. Do their efficiencies outweigh the higher costs of power and cooling? And is there a way to design or remodel your data center to ensure that infrastructure and utility costs don't outstrip operational savings?

There's no simple answer. Like a car whose mileage varies under different sets of driving conditions, your blade or virtual-server ROI will depend on the unique conditions in your facility. And your capital-budget allocation will determine when and whether to start fresh or just give your data center a facelift. To make the right deployment decisions, we advise clients to carefully consider the following factors:

Operational efficiencies. With new technology come new operational processes, which raise a variety of issues. For example, will the ability to deploy virtual-server images within minutes offset the process changes that are required around standardizing OS builds? How efficient will blades be

from an operational standpoint, and how will the management interface integrate with existing enterprise-management tools? Will having a storage and network switch within the blade chassis be an improvement over your bird's nest of cables, or will it add yet another layer of operational complexity? Overall, having less hardware, fewer cables, and less platform dependency can translate into increased operational efficiencies.

Power and cooling. It's a fact that blade-server chassis consume more energy than their single-unit 1U predecessors. Although mainframes were power-hungry, their sheer size let you spread electricity over a large expanse of floor space, yielding a low overall watts-per-square-foot power requirement. Now, depending upon size and business requirements, data centers typically demand 75 to 150 watts per square foot, compared with the 30 to 50 that mainframe facilities consumed. And consumption doesn't stop with processors: Every additional watt per square foot in computing power can substantially increase your energy needs for cooling.

Not only do blade servers drain more power than mainframes, they also emit more heat, a problem that multicore processors exacerbate. Elaborate heat-sink attachments draw heat away from the processors' core, raising the temperature of the data center. And CPUs will stay hot whether they're busy or idle.

Mainframe-oriented data centers required less cooling, since the heat output relative to the floor space occupied was small. Moreover, many mainframes in use during the late 1980s and early 1990s were water-cooled, further reducing the ambient cooling demand within the data center. The ambient cooling level of five years ago—not to mention 15 to 20 years ago—is insufficient to meet current needs.

Poor ventilation only makes matters worse. Low-raised flooring—12 inches or less—has led to drop-from-the-ceiling cable schemes. Shallow and/or congested raised-floor cavities interfere with air distribution. Low ceilings prevent warmer air from rising above IT equipment. And many facilities lack adequate hot/cold aisles or appropriately placed air-conditioning units.

A data center with several racks of blade servers requires a significant cooling infrastructure. IT equipment and cooling must be layered within each rack to dissipate heat at its point of origin.

Although water and electricity were never a good match from a safety standpoint, water cooling is making a comeback. It's much more efficient than air cooling, as it cools within the rack rather than outside of it. HP, for example, has announced water-cooled, fanless game boxes. Meanwhile, IT manufacturers are starting to design "green" technologies—think Energy Star ratings on kitchen appliances.

This will move the focus from processing capability to efficient power usage and cooling design, with the aim of substantially reducing power consumption and heat output over the next five years.

AMD, for example, has already launched a green campaign: less power, less heat, less cooling required. And in the state of California, power companies are offering an energy credit for virtualization, which decreases the server footprint and power consumption.

Floor space. Blade servers won't necessarily give you more room; in fact, you could end up with less. For one thing, you'll need to accommodate additional power distribution and backup, as well as cooling equipment. You'll also have to widen your aisles for purposes of keeping the blade-server racks far enough apart to allow for cooling and to prevent heat from concentrating in any one area.

Some data centers are building vertically, since air space is cheaper than floor space. If you go this route, take care not to overload cabinets, as they can become very difficult to cool, especially if they're close together.

Server and application strategy. Any virtualization or blade-server deployment should be part of an overall plan for your server and application infrastructure. Determine which applications are candidates for virtualization or blades, based on their performance requirements. Apps that are CPU-, memory-, or network-intensive may work best in a standalone environment.

Be aware of who's defining application-performance requirements; application owners, developers, and vendors tend to overestimate them. Take vendor guidelines with a grain of salt, using your own processes and tools to verify them. The x86 performance-management tools market has matured significantly over the past few years; today's tools can find hardware and software interdependencies and performance bottlenecks.

Blade servers will challenge an antiquated infrastructure. Before deploying blades, a thorough assessment of your projected IT needs and your current power, cooling, and floor capacity will let you know if you have to upgrade your facility or if you should build a new data center altogether.

Wholesale replacements—once known as "rip and replace" strategies—are an expensive proposition. But with IT infrastructures that are many generations beyond the mainframe for which most data centers were built, data centers are due for a significant brick-and-mortar realignment with the hardware they're supporting.

For example, the efficiency of a computer-room air conditioner (CRAC) declines 0.5% to 1% per year, assuming above-average maintenance. That means a 20-year-old CRAC operates 10% to 20% less efficiently than a new

unit. Combine that with the increase in cooling that blades demand, and the need for a CRAC upgrade becomes clear.

You'll face some major inconveniences when trying to replace the building components of an operational data center. For instance, aside from dealing with the general construction mess, you'll have to provide sufficient cooling while you're swapping out an old air conditioner for a new one.

The data center's growing centrality to doing business may force the issue of more capital investment: Customers and business partners are frequently requesting a tour of data-center facilities as part of the discovery process to make sure the infrastructure can support their needs well into the future. While it's difficult to decide to build a new \$15 million facility, the demands of customers and business partner are a convincing argument in favor of the investment.

Whatever course you take, brace yourself for Moore's Law. Floor-space needs alone grow 7% to 8% annually. When planning, request a capital budget spanning seven to 10 years so you don't have to go back to the well asking for more too soon.

Organizations that didn't plan an infrastructure for blade servers over the past five years will have power and cooling challenges if they attempt to use these technologies now. Virtualization may alleviate the situation by allowing you to consolidate servers onto fewer pieces of hardware and to better utilize your computing capacity. Before you deploy, make sure you thoroughly assess your current infrastructure's capacity and performance, and closely examine the goals of your overall data-center strategy.

Published in Information Week, July 19, 2007.

James E. Geis is director of integrated solutions development for Forsythe. Geis developed Forsythe's unique information management framework—the roadmap Forsythe uses for information management, storage and server consulting engagements. He currently is focused on developing solutions to address the impact of virtualization on computing technology and the data center.

