AMD WHITE PAPER

AMD Smarter Choice

VIRTUALIZING SERVER WORKLOADS

LOOKING BEYOND CURRENT ASSUMPTIONS

Advances in hardware virtualization technology are making it possible to virtualize an ever-wider range of x86-based server workloads. This has moved server virtualization out of its traditional domain of test and development environments and firmly into the sphere of mission-critical enterprise applications, making it imperative to expand the definition of workloads that are suitable for virtualization.

It is no longer enough to consider some workloads (file and print sharing, Web servers, and others) as suitable for virtualization while categorically ruling out virtualizing others (such as databases and e-mail servers). Every production server workload has distinct performance characteristics in storage, processing power, and memory requirements that affect suitability for virtualization. Moreover, different workloads frequently run side by side in a single organization on the same hardware— rather than the suitability to virtualization of the individual, constituent workloads, it is the aggregate of these myriad workloads running together that determines the bottlenecks to server virtualization. Understanding and taking into consideration the performance characteristics of each of the workloads, as well as the workloads taken as a collective whole, can help guide the choice of a suitable hardware platform.

Additionally, hardware enhancements have broadened the definition of what is possible with virtualization. Advances such as hardware-assisted virtualization, multi-core processors, support for faster and larger amounts of memory, input/output (I/O) improvements, and others have greatly expanded workload functionality in virtual machines.

In this white paper, we discuss the potential bottlenecks that organizations typically encounter: high memory utilization, high processor utilization, and high I/O traffic. We look at the performance characteristics of server workloads that can be successfully virtualized, and we discuss how an awareness of the performance characteristics of a particular workload can help inform an intelligent virtualization strategy. We also examine the improvements in virtualization hardware that are making it possible to virtualize an increasingly wide range of workloads.

AMD
Smarter Choice

TABLE OF CONTENTS

INTRODUCTION

Organizations face the daily challenges of having to do more with less: reduce downtime, respond quickly to new initiatives, and increase performance, all while keeping costs down. Many have found server virtualization to be an effective approach—it offers a proven way to increase equipment utilization, reduce management and administrative costs, improve the availability of services and resources, and provide a flexible infrastructure that can quickly be adapted to the changing needs of the business.

WHY VIRTUALIZE?

The traditional "one-application-per-server" deployment philosophy often leads to the use of multiple servers, many of them not fully utilized. The average server utilization in an enterprise environment can be very low, leaving much of the available capacity unused. Many customers report CPU utilization rates in the 10–15 percent range: even 5 to 10 percent utilization rates are not uncommon.

Virtualization technology not only addresses server underutilization, but can also provide additional benefits such as improved manageability and a reduction of power and cooling costs.

With virtualization, multiple workloads running on different physical machines can be consolidated onto individual virtual machines (VMs) hosted on a single physical server, harnessing the unused computing power. Reducing the number of physical servers can reduce capital costs, data center complexity, and administrative costs. Fewer servers can also reduce the IT infrastructure footprint and the associated utility costs from power and cooling.

Virtualization provides the execution isolation and service partitions that are desirable in many usage scenarios. Additionally, server virtualization can enhance business continuity strategies. Virtual machines are inherently portable, so workloads can be transferred to other physical servers during maintenance or unplanned equipment or application failures.

Virtualization also enables business agility. Server infrastructure can be quickly modified to meet changing needs or help address new business opportunities. This can be critical when workloads, such as Web-based applications, are unpredictable.

SUITABLE WORKLOADS

Some workloads are natural candidates for virtualization. General application servers, for example, are easy to virtualize because they are usually not unique—no specialized architectures are required and no special requirements need to be met. Web servers are another prime virtualization candidate: They are normally low in resource utilization and easy to move around, though they may require more planning than application servers.

Other workloads, however, pose more of a challenge, particularly those with high memory and disk I/O needs. Until recently, organizations with these workloads tended to avoid virtualization.

Traditional thinking was that some workloads just could not be effectively virtualized. This line of thinking, however, does not take into account the fact that workloads, even of the same type, can vary greatly. One database, for example, can differ significantly in size and performance requirements from another; this is true for every type of server workload. Server workloads can be characterized by storage, processing power, and memory requirements; it is these characteristics that inform suitability for virtualization, and not the workload type.

If different workloads are running side by side in a single organization on the same hardware, the amalgamation of these workloads running together on the same server determines the bottlenecks to server virtualization, rather than the characteristics of the individual, constituent workloads. Understanding and considering the performance characteristics of the individual workloads, as well as those of the workloads taken as a whole, determines suitability for virtualization and can help guide the choice of a suitable virtualization strategy.

Additionally, advances in hardware platforms are now making it possible for organizations with a wider range of server workloads to reap the benefits of virtualization. Many server workloads that were previously difficult or even impossible to virtualize can now run successfully on virtual machines. These advances include hardware-assisted virtualization—enhancements to the hardware platform that improve the performance of a virtual environment—as well as multi-core processors, support for faster memory in larger quantities, I/O improvements, and others.

These hardware enhancements have greatly expanded workload functionality in VMs, which in turn is causing organizations to rethink their virtualization strategies. By becoming aware of what is available and understanding the performance characteristics of the server workloads, IT administrators or technical decision makers can evaluate the suitability of their environment for virtualization.

CONSIDERATIONS FOR VIRTUALIZING SERVER WORKLOADS

In the past, typical IT practice followed the "one application, one server" approach: One piece of hardware and an operating system were dedicated to each significant application. This approach is simple—since applications typically require underlying software (such as relational database management systems, application servers, and Web servers), it is often less complicated to manage the application if everything is installed on a single, dedicated computer. This approach also makes it easier to allocate enough processing power and room to grow. Finally, assigning a single application per server ensures adequate isolation of applications. The default practice in many organizations is to keep important applications on individual servers so that if an application crashes and brings down the operating system, other applications are not also affected.

This reasoning is no longer appropriate for today's hardware environment. Servers are now so powerful that running a single application on a server is wasteful of processing power and leads to unnecessary server sprawl. A proliferation of servers creates unintended consequences, such as management complexities, overworked IT departments, and data centers that operate near the limit of their physical capacity. When businesses do not fully utilize the functionality of their servers, the result can be slower access to data, greater downtime risk, and increased operating costs.

Virtualization can use the power of today's servers more efficiently; it is uniquely suited to address the many issues that organizations commonly face.

COMMON BOTTLENECKS

Virtualization uses emulation to create a series of virtual machines that operate as separate hardware devices, but are in fact running on a single system; thus, a single PC can run multiple operating systems (or multiple instances of the same operating system) at the same time. With some configurations, this demands much more from the system than it was initially designed for and can lead to performance degradation: situations where one element is constraining the overall system performance.

HIGH CPU USAGE

Until recently, the main constraint with virtualization has been performance, or processing power of the host computer. Running virtual machines can provide significant advantages, but the associated overhead can also consume a large portion of the CPU processing cycles, reducing workload performance. Servers have traditionally been designed so that the CPU runs one operating system, with one set of applications and one set of users. With virtualization, that same CPU is being tasked to handle multiple operating systems, multiple sets of applications, and multiple sets of users. Historically, conventional wisdom has shown that if a CPU-intensive task is running on a virtual machine—for example, indexing a multi-million record relational database—there will be performance degradation across all of the servers in the environment.

In truth, however, CPU needs are often the easiest to resolve. Recent-generation computers have enough available processing power to ensure that the processor is less likely to be the performance bottleneck in the system.

Prior to the introduction of hardware-assisted virtualization, x86 processors were designed for the "one operating system, one server" model; they assumed that a single operating system was running on a single physical server and had access to all the resources on that server. The latest generation of processors, however, is designed specifically to support the virtualization model, making it possible for server workloads with high performance requirements to be virtualized. The processor assists virtualization by offloading some of the processing from the software to the hardware, improving the efficiency of the implementation. An example of this is AMD Virtualization™ technology, or AMD-V™.

Dual-core and quad-core processors are widely available from AMD and other manufacturers, and soon there will be multi-core processors with even higher processor density; these processors can help a data center consume less power and require less cooling, providing an excellent base for virtualization.

HIGH MEMORY REQUIREMENTS

Though it can be easy to focus on processing power, from a cost perspective, memory and I/O can be much more expensive to provision. Multi-core CPUs can reduce the cost per processor core, but memory cost still remains at a premium. In fact, memory is often the critical resource for virtualization; it can be the most common bottleneck for virtualization performance and is often the most difficult to accommodate in hardware.

Traditional thinking is that the amount of memory space that is required to virtualize an environment can be prohibitive. A virtual PC running a legacy version of Microsoft® Windows NT® 4.0 as a guest operating system requires about 1.5 GB of disk space and 128 MB of RAM. For the VM, roughly 2 GB of disk space and almost 200 MB of RAM must be added just to get the system up and running.

Modern servers come with up to 2 GB or more of memory—more than enough for most application loads, especially in a "one application, one server" mode. With virtualization, however, a server may support 10 or more VMs—so adding 1 GB per VM (a typical amount to add) means that far more memory is needed for adequate performance.

Managing all of this memory can also demand significant resources. The operating systems must maintain page tables to translate the virtual memory page into physical memory addresses. Until recently, the guest operating system running on a VM could only see shadow page tables—page tables that ran on an emulated memory management unit (MMU)—and had no access to the real page tables. The real page tables, managed by the virtual machine manager (VMM), ran on the real MMU. Modifying and adjusting the shadow page tables is extremely CPU-intensive and often results in significant overhead. In fact, with memory-intensive applications, memory management accounts for the largest part of the virtualization performance penalty.

In response, AMD recently introduced Rapid Virtualization Indexing (RVI) as part of the AMD-V in Quad-Core AMD Opteron™ processors; this feature helps to eliminate the need to use shadow page tables and improves the performance of many memory-intensive virtualized applications.

Common techniques that support VM workloads are also memory-intensive. Mechanisms for managing memory include a ballooning technique that reclaims pages that are considered least valuable to the VM operating system, as well as content-based page sharing and hot I/O page remapping, both of which can help eliminate redundancy and reduce copying overheads.

Advances in hardware technology have also helped to make it possible for workloads with high memory requirements to be virtualized. AMD multi-core processors with built-in integrated memory controllers are designed specifically to maximize the performance of memory-intensive virtualization environments. Virtualization extensions such as the AMD tagged translation lookaside buffer (tagged TLB) and Rapid Virtualization Indexing help improve the performance associated with managing memory of the different guest operating systems running on a single physical server—enabling more efficient switching between virtual machines by maintaining mappings to the individual memory space for each VM.

HIGH INPUT/OUTPUT TRAFFIC

Computer systems generate significant amounts of data, and adding virtualization compounds the situation: Therefore, I/O is a significant consideration in virtualized environments. While it is possible to add CPUs or upgrade to multi-core CPUs if a virtualized server needs additional CPU power, it is more difficult to upgrade the memory bandwidth, the storage Hardware Bus Adapters (HBA), and the chipsets, all of which are generally shared by all of the virtual machines.

In software virtualization, the hypervisor software traps the machine operations that the operating system uses to perform I/O operations or to read or modify the system's status. The hypervisor then emulates these operations in software, and returns status codes that are consistent with what the real hardware would have delivered in a native environment. This instruction trapping and emulation is necessary—the memory would likely be corrupted if the operating system tried to instruct the hardware to perform direct memory access (DMA) because the hardware cannot distinguish the difference between virtual address mapping (used by the guest operating system) and physical address mapping—but it can reduce the overall system performance in I/O intensive environments. For this reason, the overhead can be much higher with I/O-intensive workloads than with those that are compute-intensive or memory-intensive.

Workloads that have a high amount of I/O traffic have thus not been considered to be good candidates for virtualization traditionally. Indeed, throughput is often the primary limiting factor when implementing a virtualization solution.

I/O constraints also impact the design of the back-end storage. When virtualizing in a production data center, networked storage is a critical element to consider. In development and test environments, virtualized applications are typically run with local disk or direct-attached RAID storage. In production data centers, however, VMs need to work with enterprise-class storage area network (SAN) or network attached storage (NAS) that is shared across a range of applications and workloads. Since storage relies on I/O, it is important to ensure that the I/O workloads required by the server domain can be handled by all of the elements of the storage domain, including the HBA, the storage fabric, and the storage array.

Performance issues in virtualized server environments are often the result of a mismatch between the front-end workloads and the back-end storage; contention for shared storage resources can cause I/O bottlenecks that lead to queuing backlogs and poor end-to-end response time. It may be possible to add additional VMs to a given server, but this may overload the storage layer. Enterprise storage is therefore increasingly virtualized along with servers and acts as a pooled, shared resource.

I/O memory management units (IOMMUs) help reduce I/O overhead by re-mapping the addresses accessed by the hardware according to the same (or a compatible) translation table used by the virtual machine guest—thus enabling a wider range of high I/O server workloads to be virtualized. The hypervisor dedicates a portion of the system memory to a particular guest VM when it is initiated. That VM can then directly access the memory without going through the virtualization software. The overhead is thus restricted to the startup phase of the guest VM, rather than being imposed for every memory access operation. In addition, I/O virtualization technology enables secure partitioning at the peripheral component interconnect bridge level (computer bus for attaching peripheral devices to a computer motherboard); this allows enforcement of device ownership at the very lowest levels of the platform.

TYPICAL WORKLOADS

The performance characteristics of a particular workload, in terms of the three bottlenecks, are critical to consider when embarking on a virtualization implementation. While it is true that some workloads are easier to virtualize than others, advances in hardware have opened the door for many workloads that had previously been considered difficult, or even impossible, to virtualize. Understanding how server workloads behave lets organizations move beyond conventional thinking and take advantage of these new advances.

To help evaluate the performance characteristics of a workload, it is important to consider the characteristics of typical server workloads: databases, Web servers, file and print servers, terminal servers, desktops, and e-mail servers. Though each has its unique challenges, all are being successfully virtualized today.

DATABASE WORKLOADS

Database virtualization can reduce the cost of maintaining dozens of custom data marts, enabling movement of older databases to commodity hardware (or letting them be retired altogether), and reducing the costly and time-consuming process of copying data. Database virtualization also provides more flexibility in deployment and rapid response to change. Databases, however, have some unique properties that can make them more complex to virtualize. As a group, they have been traditionally considered poor candidates for virtualization.

The issue with virtualizing databases within a server virtualization platform has been the perceived I/O bottleneck that comes along with virtualization. Databases can be characterized as having:

- *Large memory:* Databases use very large amounts of memory to cache their storage. A large cache is one of the most important performance criteria for databases, since it can significantly reduce the physical I/O.
- *High performance block I/O:* Databases read and write their data in fixed, block-sized chunks. The I/O blocks are typically small, and operate at a very high rate on a small number of files or devices.
- *High throughput:* Databases often have a large number of concurrent users, giving them natural parallelism and making them ideally suited to take advantage of systems with multiple logical or physical processors.

Brandon Worrell, lead solutions architect at Solutions-II, a national IBM Premier Business Partner focused on system architecture, deployment, server and storage

management, business continuance, and server consolidation initiatives, describes the two main types of databases that Solutions-II encounters during virtualization projects: data warehouses and online transaction processing (OLTP). According to Worrell, each has very different I/O characteristics, and these characteristics generally have the greatest effects on the architecture of the back-end storage.

- Data warehouses are characterized by large-block sequential transfers, and therefore bandwidth is the gating factor (measured in MB/sec).
- OLTP databases are characterized by small-block random transfers, and therefore the ability of the storage to quickly provide cache read-misses is of primary importance (measured in IOs/sec or transactions/sec).

WEB SERVER WORKLOADS

Web servers are generally easy to virtualize. They tend to be characterized by CPU utilization that is generally low but spikes during peak periods, and CPU utilization is perhaps the easiest resource utilization constraint to mitigate. A Web server's performance is thus most affected by the performance of the CPU and memory, especially if there is any server-side scripting.

In virtualizing Web servers, it is important to differentiate between static and dynamic Web sites. Static (HTML only) Web sites are the easiest to virtualize, since they usually rely only on the Web server. Dynamic Web sites, however, also typically require a database server. Either way, the Web server should be assigned the greatest amount of CPU and memory resources as possible. With a dynamic Web site, there is an additional challenge: Database server performance is disk-bound because of the large amount of disk I/O on a database server.

Additionally, as Worrell of Solutions-II points out, networking and security must be considered. If the virtual Web server will be public-facing, it will most likely be located in a perimeter network or demilitarized zone (DMZ) (whereas the database server is typically inside the firewall). To connect to the separate database server on a dynamic Web site, ports can be opened on the local area network (LAN) side of the DMZ, or the database server connection on a virtual LAN (VLAN) can be segmented and connected. If the virtual Web servers are on the same physical host machine as other applications, it is possible to segment the Web servers into their own VLAN by assigning a dedicated physical network card on the host machine to the Web servers. This network card should be connected to the appropriate VLAN segment. In the case of VMware ESX Server, a user can also create a new VLAN that is internal on the ESX Server by segmenting the virtual switch into a VLAN. Note that if servers are deployed across multiple VLANs in a shared hosted environment, "inter-VLAN" VM traffic has to traverse many layers of network infrastructure, including an external firewall, to reach its destination, even if the VMs are on the same physical server and attached to the same virtual switch. A Web server request that should take only a few hops to get to its destination becomes much more complex when it involves VLANs and external firewalls in a virtualized environment.

TERMINAL SERVERS

Terminal Servers can be challenging to virtualize, as they tend to have very high memory and disk I/O needs. Most applications used on Terminal Servers are constantly writing to disk and loading data into memory. One strategy to circumvent this challenge is to load the user profiles on a physical server and the applications on separate volumes—this keeps the memory utilization and disk I/O more moderate, as the user profile data uses the network interface card (NIC) and the application data uses the HBA for the data store.

FILE AND PRINT SERVERS

File and print servers tend to use a lot of disk space because most applications are not running locally on the server. Though they are storing data, memory and CPU usage tend to be low. However, if a user is running large file servers as well as virus scan software, there can be high CPU, memory, and disk I/O. Cost tends to be a consideration with file and print servers; if expensive SAN disks are used to store data that is not often used, it may not make sense to virtualize that data.

VIRTUALIZED DESKTOPS

Desktops can be virtualized and centrally managed. In a virtualized desktop solution, multiple virtual machines run on a standard, single-user desktop PC operating system hosted on a central server. A wisely architected desktop virtualization solution can provide great benefits—including cost, security, and manageability benefits—but users' needs must be considered. An installation for a developer running a server with a Microsoft® SQL Server® database will look very different from one constructed for a general office worker who uses Microsoft Office, Web applications, and possibly a mainframe emulator.

One challenge with virtualized desktops has been providing a rich graphical experience to the user. This is mainly because virtualizing the Graphics Processing Unit (GPU) is very complex and incurs significant overhead. Even if these problems are sufficiently addressed, delivering a rich graphical experience can consume significant network bandwidth. Applications such as computer-aided design (CAD) tools and computer games, therefore, may not be suitable in virtualized desktop environments.

For additional functionality, a connection broker can be used; it connects the client access device on a user's desk to the back-end, central server resources (the target resources). A connection broker, also called an infrastructure access package, can perform a variety of tasks depending on the version selected. At the basic level, the connection broker directs incoming requests to an available hosted desktop. In some cases, a connection broker can integrate with Lightweight Directory Access Protocol (LDAP) or Active Directory® to authenticate users. The connection broker can use a predefined policy or group membership to direct (or assign) users to a hosted desktop, control the state of a hosted desktop instance (power it on or off, for example), or track the connection status of a hosted desktop. Some connection brokers also offer secure sockets layer (SSL) or IP security (IPsec) functionality for secure virtual private network (VPN) access.

E-MAIL WORKLOADS

As with other workloads, there are performance considerations, support limitations, and deployment issues that have to be taken into account before virtualizing any part of an e-mail server within a production environment. For e-mail servers, disk I/O is the main consideration.

For example, it is generally a straightforward task to virtualize front-end servers running Microsoft® Exchange Server (or client access servers)—even under stress, a VM with a single CPU and between 512 MB and 1 GB of RAM should be sufficient for a server providing Microsoft® Outlook® Web Access (OWA). However, it is difficult to virtualize the back-end servers running Exchange Server (mailbox servers in Microsoft® Exchange Server 2007) because of the extremely high amount of disk I/O that Exchange Server generates—virtual disk files simply have a hard time keeping up.

COMBINING WORKLOADS

Though it is important to understand the characteristics of the individual workloads when considering virtualization, different workloads frequently run side by side in a single organization on the same hardware, and it is also critical to take the aggregate effects into account; it is the aggregate of the workloads running together that determines the bottlenecks to server virtualization, rather than the suitability of the individual constituents to virtualization. Overlooking the aggregate effects can lead to unnecessarily limiting the potential gain in efficiency or failing to leave enough headroom to cushion peak demands on the infrastructure.

To help determine if an entire combined workload environment is suitable for virtualization, consider the following measure of performance characteristics:

- **Aggregate utilization** is measured by normalizing the workload curves of all physical servers against their overall "power" (typically obtained using benchmarks) and summing them to obtain a weighted average. The per-hour, time-of-day curves can also be normalized and summed to give a view of the aggregate workload pattern over time, which then shows the distribution of resource demand in the target environment.

  At Solutions-II, describes Brandon Worrell, they utilize capacity planning tools, such as VMware Capacity Planner and PlateSpin PowerRecon, to measure workloads over time to determine the workload averages and peaks. This lets them determine the length and time of peaks and plan appropriately. "We also count on VMware DRS in virtualization installations to automatically adjust the size of the VMs as unexpected peaks occur," says Worrell.

- With **what-if analysis**, it is possible to assess the various combinations of workload patterns to determine the optimal stacking function for the environment. This analysis involves normalizing the workloads against the relative powers of the source and the target servers and then stacking specific sets of workloads onto target systems to determine which combinations fit best.

  Again, Solutions-II uses capacity planning tools to help inform their architectural decisions. These tools are not perfect, though they continue to improve in accuracy. According to Worrell, "at Solutions-II, we have found that we must also rely on experience we have gained in the field to supplement the information we get from the capacity planning tools."

- **Overhead modeling** enables evaluation of the overhead related to I/O and other operational issues. Compensating for this overhead is important to ensure sufficient capacity is allocated in reserve to sustain target service levels, and any efficiencies gained in the process (for example, elimination of multiple backup devices) should be accounted for to fully optimize the resulting environment.

  Worrell adds that architecting the back-end storage is extremely important in combined virtualized environments. "For example, even if a VM generally has low I/O access that is primarily sequential, combining 50 of these VMs onto a virtual farm turns that sequential access into random access because of the sheer number of computers," says Worrell.

Virtualization implementers such as Solutions-II have gained enough experience with virtualization over recent years to have developed strategies to avoid some of the common bottlenecks when combining workloads.

For example, a production environment with high I/O requirements and CPU-intensive applications can divide each of the applications onto multiple VMs. A single VM running an I/O-intensive application and a CPU-intensive application together will run more efficiently than two separate VMs (one with a I/O-intensive application and one with a CPU-intensive application). Thus, it is better to have a Web server and a database on the same VM, and then have another Web server and database on another VM. In this way, the resources of each VM are more fully utilized, rather than just the I/O of one VM and the CPU cycles of the other one.

Worrell describes another example: When Solutions-II designs storage and decides which VMs to combine into a virtual machine file system (VMFS), they have found that it is best to combine sequential low-access VMs with a number of high-access random VMs. If tiered storage is used, then it is especially important to model the storage characteristics of the VMs.

ADDRESSING THE CHALLENGES

Though the CPU, I/O, and memory bottlenecks are very real, it is important to look beyond the perceived limitations. Advances in hardware technology, as well as creative configuration, have opened the door for the virtualization of many non-traditional server workloads.

The use of virtualization in production has dramatically increased because of the improved capabilities and lower cost of hardware. Quad-core processors are widely available today at less cost than older single-core processors. Memory is much denser—servers commonly allow 64 GB, 128 GB, or more. The speed of I/O has increased as well. These advances all contribute to the ever-expanding range of virtualized server workloads.

HARDWARE-ASSISTED VIRTUALIZATION

Operating systems do not expect to have to share physical resources. However, sharing resources is one of the fundamental advantages of virtual machines. As discussed earlier, memory and processing requirements for virtualization can be high. Hardware virtualization can address the enormous overhead imposed by software virtualization by moving many of the computational tasks associated with platform management to the hardware, thus removing a layer of abstraction and letting the CPU take on some of the "heavy lifting." Emulating at the hardware level is much faster than software emulation, so encoding the capability for virtualization at the hardware level helps minimize the CPU, memory, and I/O bottlenecks encountered when virtualizing in a production environment.

In the traditional "one computer, one operating system" computing model, the operating system is able to alter, completely unchecked, the state of the CPU, the chipset, and the peripherals. A virtualized system is different: It must be able to ensure that an operating system cannot alter the system state in a way that would prevent the computer from being shared among multiple operating systems.

Software alters the system state through the execution of privileged instructions. One of the most difficult tasks that software emulation must handle is the identification and redirection of these instructions. Hardware-assisted virtualization steps in to provide the CPU with the capability to intercept and redirect requests to alter the state of the system.

When an operating system runs on a processor that supports hardware-assisted virtualization, any privileged operation can be intercepted before completion and directed to an entry point set up by the virtualization layer, which holds and grants the privilege to alter the system state. The privileged instruction intercepts are built into the CPUs; they incorporate the saving and restoring of extended system state into new instructions.

Some of the first examples of hardware-assisted virtualization have come from AMD. AMD provides a suite of hardware-assisted virtualization technologies, known collectively as AMD Virtualization™ (AMD-V™) technology. AMD-V simplifies the processes within the virtualization layer that are associated with trapping and emulating I/O operations and status instructions executed within the guest operating system. By decreasing, and in some cases eliminating, the virtualization overhead associated with processor operations, performance is improved.

MULTI-CORE PROCESSORS

Virtual machines require physical resources to be scheduled for them—if a VM requires two virtual CPUs, the hypervisor must wait for two CPUs (or cores) to be available in order for the VM to run. Multi-core CPUs and the additional parallelism they provide have opened the door to a wider range of multi-virtual CPU workloads than ever before. The VMware ESX Server 3 provides four-way Virtual SMP (vSMP), which lets a single VM use up to four physical processors simultaneously. This simultaneous processing capability makes it possible for CPU-intensive applications such as databases and messaging servers to be virtualized.

HARDWARE PAGE TABLES VIRTUALIZATION

Memory virtualization, including the partitioning and allocation of available physical memory among the VMs, can be assisted by hardware page table virtualization. With memory virtualization, VMs see what appears to be a contiguous address space, but it is actually not contiguous within the underlying physical memory. The guest operating system stores the mapping between the virtual and physical memory addresses in page tables, and because they do not have native direct access to the physical system memory, the virtual memory manager (VMM) must perform another level of memory virtualization in order to simultaneously accommodate multiple VMs—the mapping between the physical memory and the page tables in the guest operating systems must be performed within the VMM. Rapid Virtualization Indexing (RVI) helps to accelerate the additional layer of memory translation that is required.

To measure the performance improvement of Web servers with RVI, AnandTech, an online source for hardware analysis and industry news, recently ran a series of benchmarking tests. Benchmarks were run on a dual-socket AMD Opteron™ processor-based system (eight cores at 2.3 GHz). Four VMs were run with two virtual CPUs linked to two physical cores. Two Web servers (one running Oracle OLPT and one running DSS MySQL) were run in parallel on the server. Each VM had 4 GB of RAM and ran Windows Server® 2003 R2. RVI was enabled and disabled in the kernel parameters of Xen 3.2.0 (SUSE SLES 10 SP2). RVI improved performance by 31% for the PHP Website on Windows Server 2003 R2, and 7% on the Oracle Swingbench OLPT test. [1]

I/O VIRTUALIZATION

server I/O uses physical interfaces with fixed identities that are mapped to storage and network resources. Because these mappings are time-consuming to change, applications become locked to specific devices. This has an impact on server performance and resource utilization. The mappings may also be managed by multiple server, storage, and networking teams, so any change may require coordinating multiple groups. A simple application move from one server may require weeks to execute. I/O virtualization addresses these issues by letting IT administrators reconfigure, re-map, and re-cable resources without affecting servers, storage, and networking gear. The I/O virtualization appliance helps reduce server connectivity bottlenecks by replacing cabling and network and storage interfaces with virtual resources.

Three specifications help enable virtualization solutions to tackle I/O-intensive workloads by removing performance bottlenecks in both software and hardware virtualization components:

- **Address Translation Service (ATS)**
  ATS optimizes performance between an I/O device and the platform's input/output memory management unit. By using translated addresses, cache pressures can be reduced on the IOMMU; this reduces memory bus consumption and contributes to optimal performance.

- **Single-Root IOV (SR-IOV)**
  SR-IOV lets multiple guest operating systems simultaneously access an I/O device without having to trap to the hypervisor on the main data path. Direct hardware access can significantly improve system performance, reduce system power consumption, and lead to greater cost savings.

1 *For details see Johan De Gelas, The very first independent Nested Paging Virtualization tests, http://www.anandtech.com/weblog/showpost.aspx?i=467. Configuration: 2 Quad-Core AMD Opteron™ Model 8356 processors (2.3 GHz), 4GB RAM, Microsoft® Windows [Server?]® 2003 R2.*

- **Multi-Root IOV (MR-IOV)**
  MR-IOV lets either PCI Express® (PCIe®) or SR-IOV I/O devices be accessed through a shared PCIe fabric. This sharing makes it possible for fewer I/O devices to be provisioned, reducing system power consumption and hardware provisioning costs.

CAPACITY PLANNING TOOLS

As discussed, successful virtualization initiatives cannot be undertaken without a thorough understanding of the server workloads. Previously, managers relied on best guesses and intuition to identify underutilized or under-protected servers, and to allocate sufficient resources for current and future needs. New capacity planning tools can quantify complex server consolidation, disaster recovery, capacity planning, and other data center initiatives by remotely discovering software and services inventory across the environment and then analyzing key workload utilization metrics in order to develop optimal virtualization plans.

While some capacity planning applications only consider average workloads when determining whether they can be consolidated, others now consider peak workloads and when they occur. For example, the aggregated average workload of two servers may exceed the capabilities of the host server, but their workloads may peak at different times. This suggests that it may in fact be possible to consolidate them on the same virtualization host.

INDUSTRY EXAMPLES

Organizations are taking a second look at production environments that were previously considered impractical, or even impossible, to virtualize successfully. Of course, production environments can be complex—and can encompass a multitude of different server workloads.

David St. Clair, a consultant with Toronto-based InFront Consulting, gives examples from his practice in which the common bottlenecks were considered and addressed. These examples demonstrate that looking at the performance characteristics of server workloads, rather than relying on assumptions that may be outdated, can open the door to a variety of new implementations—though challenges may still, of course, exist.

One InFront client brought in large four-way servers (four processors with 32 GBs of RAM and SAN data stores) and undertook a physical-to-virtual (P2V) migration of about 200 servers consisting of Web, standalone, and mid-tier applications. They also virtualized one domain controller from each domain. All new applications entering production were virtualized unless they proved not to work or unless there were problems with vendor support. They also migrated entire developer environments to quad-core servers with 32 or 64 GB of RAM. In all, about 20% to 30% of production was virtualized.

The first bottleneck they encountered was the CPU usage. St. Clair found that his team was able to put certain applications (those that were not multi-threaded) that had previously been running on physical dual-processor servers onto individual virtual servers each assigned a single virtual processor; this addressed the processing issue by reducing both processor wait times and cost. I/O was also an issue and a dedicated SAN was installed to help resolve these issues. Some extremely busy servers, such as those running SQL Server, were kept on separate physical servers; others were tuned and performed well on virtual machines. Fibre Channel-attached disks also helped optimize the I/O traffic.

A second example St. Clair gives is an organization that had two environments, *.com* and *Corp*. Their .com environment was composed of a large number of development environments, all built virtually. These environments were mostly (95%) composed of Web servers. The workloads, and therefore the resources needed by the virtual machines, were very minimal: Most had 1 VCPU, 512 MB of RAM, and a 10 GB C: volume.

The *Corp* environment consisted of front-end systems running Microsoft® Exchange Server, domain controllers, Web applications, mid-tier applications, small dedicated database servers (mostly dedicated application databases with small performance requirements), file and print servers, developer workstations (running server and application-building tools), offshore Terminal Server farms (consisting of 30 Terminal Servers), and entire point-of-sale (POS) development and test environments (roughly 20 environments with three to four servers each). The plan was to virtualize the entire store environment and move from a physical-first model to a virtual-first model. All new applications would be virtualized unless it was proven that virtualization would not work or was not supported for a particular application.

The company started with small host servers—dual-processor, 8 GB RAM servers running 15 guest VMs each. All of the guest VMs were memory constrained, ran above the recommended processor utilization, and had high I/O from the large amount of paging to disk. The environment was then rebuilt with much larger host servers. Memory and processor utilization dropped to below normal ranges (half of the total memory utilized, processor utilization around 10%-20%). The RAM capacity for the virtual machine running Exchange Server was increased from 512 MB to 2 GB, and an extra virtual processor was added. This significantly reduced queue lengths, and also noticeably improved overall performance. All of the servers received a memory upgrade to at least 1 GB (some to 2 GB).

Servers running SQL Server were tested, and as long as they were small application-dedicated SQL Server versions, the I/O was perfectly acceptable. However, the larger enterprise SQL Server-based servers did encounter some bottlenecks in memory and I/O. Most of the Web applications functioned well with normal resource specifications (depending on the Web farm, the guest VMs were between 512 MB and 1 GB of RAM, single-processor servers). All performance metrics fell in or below normal utilization.

The domain controllers were tricky to virtualize, adds St. Clair. There were primary replication partners as well as standard global catalog (GC) domain controllers. The standard GC domain controllers ran well with a single processor with 1 GB of RAM, but the primary replication partners tended to run higher, with memory utilization and the disk I/O that was above normal. The NIC utilization was also much higher; using shared network resources can be a bottleneck depending on the applications running on the host and on the host architecture.

The main data center setup consisted of six host environments with dual/quad-core processors, 32 GB of RAM, and SAN volumes for all VMs. A development environment consisting of a smaller local data center with an HP C-Series Blade enclosure was also set up: six hosts with dual/quad-core processors, 16 GB of RAM, and SAN data stores.

The organization was moving away from a single guest size limit to more of a built-to-fit model, allowing large, enterprise-class servers running SQL Server or Exchange Server to run on dedicated host hardware. When two large servers running SQL Server were needed, one standard host was split into two dedicated SQL Server guests (allowing them to split the 32 GB of RAM and the quad-core processors). The I/O was carefully monitored.

Overall CPU utilization after the rebuild was roughly 10% total utilization—significant overhead was added for future growth. The overall memory utilization was below 50%, again allowing for future growth and failover. The overall I/O utilization was about 40%. As St. Clair's examples show, there are numerous considerations when virtualizing a production environment, but it is a venture well worth pursuing.

CONCLUSION

Until recently, the cost of entry to a virtualized infrastructure was very high and the applications that could be easily virtualized were relatively limited. With the advent of affordable, robust virtualization on the x86 platform coinciding with the introduction of inexpensive, high-performance, and highly reliable server hardware, virtual machine technology is now accessible to a broad audience. Virtualization is no longer limited to development and testing; research shows that more and more enterprises are now using virtualization for their production applications.

And while virtualization has moved out of its traditional domain into mainstream production, conventional thinking about its capabilities and limitations still keeps many organizations from embarking on large implementations. The suitability of workloads for virtualization in production environments is often guided by assumptions that are no longer true. Because every server workload has distinct performance characteristics, understanding and taking those characteristics into consideration is necessary to determine the server's suitability for virtualization. Improvements in virtualization hardware and other advances in virtualization technology are making it possible to virtualize an increasingly wide range of workloads.

In summary, organizations that decided against virtualization in the past because of concerns about CPU, memory, or I/O bottlenecks would do well to take another look.

APPENDIX: AMD ADVANTAGES

AMD has long been known as a leading designer and producer of microprocessors. As the use of virtualization has grown, AMD has committed to enhance the performance of virtualization workloads with architectural changes. Processors based on AMD64 technology—including AMD Opteron™, AMD Athlon™, and AMD Phenom™ processors—reflect this commitment by providing the underlying architecture that inherently aids virtualization, and adding the additional virtualization-specific capabilities needed to efficiently run multiple operating systems.

DIRECT CONNECT ARCHITECTURE

All AMD64 processors are built on Direct Connect Architecture, which helps reduce the bottlenecks inherent in 20-year-old front-side bus architectures by directly connecting CPUs, memory, and I/O for low latency and optimal memory performance.
AMD Direct Connect Architecture provides direct CPU-to-memory, CPU-to-I/O, and CPU core-to-core connections to streamline server virtualization. Quad-Core AMD Opteron processors provide enhanced memory bandwidth and CPU resources for leading-edge virtualization performance.
Components of Direct Connect Architecture include the following:

- **AMD64 technology** offers 64-bit memory addressing, which enables virtualization software to efficiently handle multiple guest operating systems and applications.

  Processors are also compatible with 32-bit x86-based operating systems and applications, allowing legacy environments to be virtualized on newer, more power-efficient servers.

- **Multi-core processing** provides the processing resources needed to effectively drive multiple virtual machines, making it possible for multiple applications to be consolidated onto one server.

- **HyperTransport™ technology** helps reduce I/O bottlenecks and provides multiprocessor scalability, both of which aid in consolidating workloads. HyperTransport technology optimizes the movement of data and the sharing of resources among VMs for greater system scalability.

- The **integrated memory controller** provides fast access to memory, meeting the demands of inherently memory-intensive virtualized environments. Since memory is "owned" by the CPUs, advanced memory handling can help increase the security of virtual machines. The integrated memory controller is designed to improve performance on memory-intensive virtualization environments through high bandwidth throughput, low latency, and scalable access to memory.

AMD VIRTUALIZATION (AMD-V) TECHNOLOGY

Supplementing the inherent benefits provided by Direct Connect Architecture, AMD also created processor improvements that specifically benefit virtualization. These improvements can be found in AMD Opteron processors and are collectively known as AMD Virtualization™ (AMD-V™) technology.
AMD-V is built on the Direct Connect Architecture foundation, which reduces overhead by allowing direct communication between guest virtual machines and the physical processor(s), and by providing enhanced memory handling.

- **Rapid Virtualization Indexing** allows virtual machines to more directly manage memory, helping to improve performance on many virtualized applications. Utilizing on-die silicon resources rather than software, Rapid Virtualization Indexing can minimize the number of hypervisor cycles needed, as well as the associated performance penalty that is commonly associated with virtualization.

  Rapid Virtualization Indexing is also designed to minimize the "world-switch time"—time spent switching from one virtual machine to another—for faster application responsiveness.

- The **Tagged Translation Look-aside Buffer (TLB)**, which is unique to AMD Opteron processors, allows for faster switching times between virtual machines by maintaining a mapping to the individual memory spaces used by the VMs. Distinguishing between the memory spaces used by each VM helps reduce memory management overhead and enhances responsiveness when switching between virtual machines.

- **AMD-V Extended Migration** is designed to enable virtualization software solutions to achieve live migration of virtual machines across the entire current range of AMD Opteron processors.

DISCLAIMER AND ATTRIBUTION

46076-A