Server virtualization technologies for x86-based HP BladeSystem and HP ProLiant servers

technology brief, 3rd edition

Abstract

Virtual machine technology is the most widely understood and adopted form of server virtualization technology today. It is a powerful technology that can reduce overall costs, decrease power consumption, footprint requirements, and improve resource utilization. This technology brief describes how HP is using virtual machine technology and expanding into other areas of server virtualization such as physical-layer abstraction and management technologies for virtualization. These include ProLiant iVirtualization, HP Virtual Connect, logical servers, and HP Insight Dynamics – VSE. This technology brief is written with the assumption that readers understand the basics of virtual machine technology.

Introduction

The term *server virtualization* refers to abstracting, or masking, a physical server resource to make it appear different logically to what it is physically. In addition, server virtualization includes the ability for an administrator to relocate and adjust the machine workload. In its broadest sense, server virtualization refers to techniques implemented using either software or hardware. However, the term *server virtualization* is commonly used as a synonym for virtual machine technology, because that is the most widely adopted form of server virtualization technology today for x86-based servers. To avoid confusion, this technology brief uses the term *machine abstraction* to refer to the broad set of technologies that abstract an entire physical server and allow its resources to be pooled and shared.[1] Machine abstraction logically abstracts and isolates the operating system (OS) and application workload from the underlying hardware, and can be implemented using software or hardware techniques.

Machine abstraction technology is important because as more server resources are abstracted, the IT organization can become more efficient and effective:

- Reduces total cost of ownership as physical resources are used to their full potential
- Increases flexibility because resources can be provisioned or moved as needed
- Increases scalability because resources can be scaled up or down based on changing workload demands
- Improves resiliency by simplifying backup, failover, and disaster recovery solutions

Machine abstraction can be performed in the software layer (virtual machines), or physical layer. HP introduced *physical-layer abstraction* with HP Virtual Connect, which abstracts the server-to-network connections. HP also introduced the concept of logical servers, allowing administrators to establish server profiles, or containers, to describe an abstracted server. Figure 1 indicates the locations where logical servers can be landed:

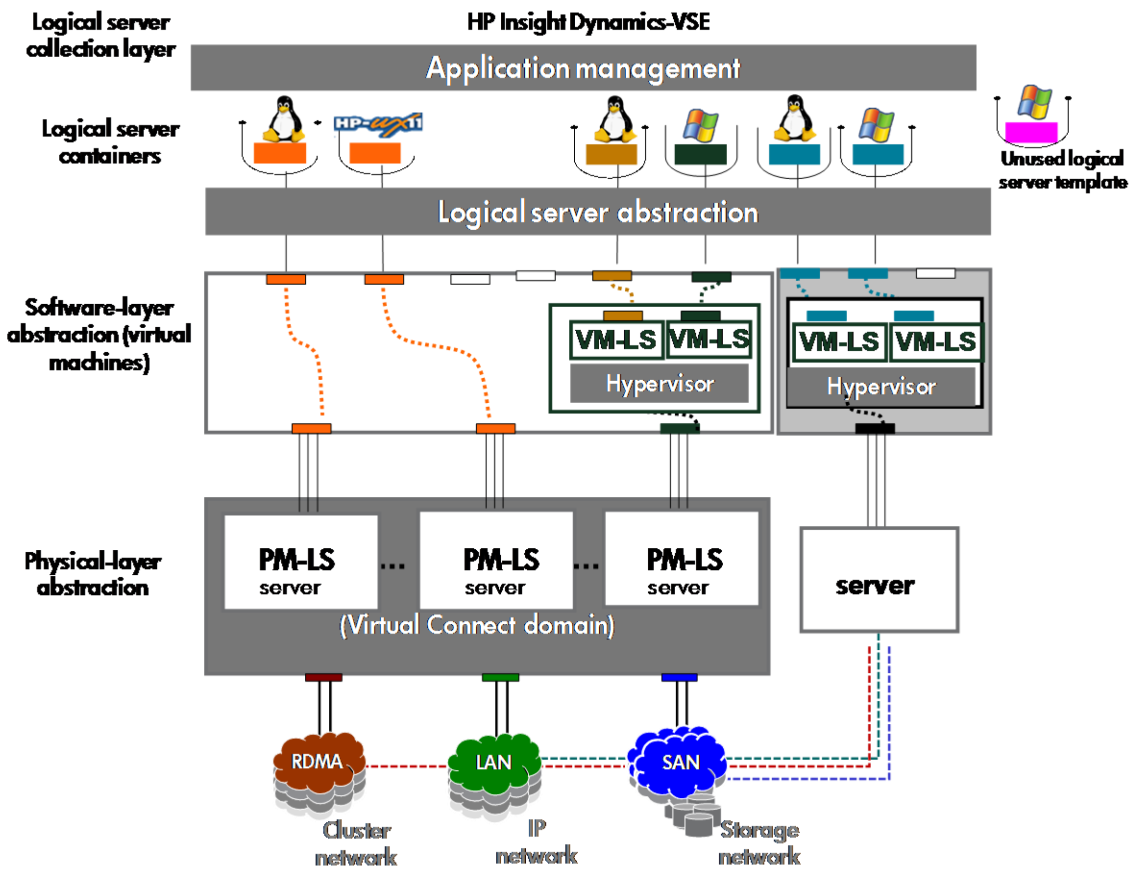
- Pool of servers such as a BladeSystem Virtual Connect domain (referred to as a physical machine-logical server, PM-LS)
- Hypervisor-based virtual machine (referred to as a virtual machine-logical server, VM-LS)

Figure 1 also shows a centralized management server (CMS) as the top-level orchestration manager for the data center. Using a CMS, server administrators have a single console and server management tool that spans multiple physical machines, gives a comprehensive view of the server infrastructure, and can act as a "collection" tool for all the machine abstractions (physical-layer or software-layer). Figure 1 shows HP Insight Dynamics – VSE as an example of a CMS.

HP is bringing Virtual Connect, virtual machine, and integrated management technologies together so IT administrators can create and manage virtual machines and physical machines in the same manner throughout the lifecycle of a server. The following sections discuss each of these technologies.

[1] Sharing may be done concurrently or sequentially.

Figure 1. HP technologies for machine abstraction include Virtual Connect, integrated hypervisors, and integrated management.

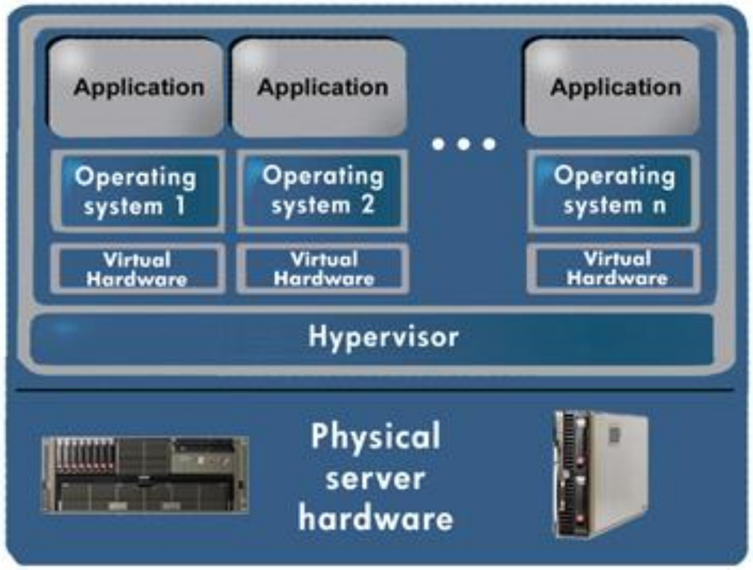


Software-layer abstraction: virtual machines

In a virtual machine environment, a software layer abstracts the physical server hardware and creates one or more virtual machine instances, each with its own virtual drives, virtual network interface controllers (NICs), virtual storage controllers, virtual processors, OS, and application(s). The software abstraction layer is typically referred to as a hypervisor or a virtual machine monitor. An OS that runs in the virtual machine instance is called a guest OS.

On a single physical platform, the software abstraction layer can create one or more virtual machines. These virtual machines share the physical resources (Figure 2). In today's data centers, servers often run single workload instances and are sometime under-utilized in off peak hours. By moving these workloads onto virtual machines, administrators can consolidate under-utilized hardware, and better utilize the server hardware while still maintaining OS and application isolation.

Figure 2. In virtual machine technology, the software abstraction layer (hypervisor) abstracts the physical-layer resources.



In x86 environments, there are several variations[2] within software-layer abstraction of the server hardware, including these general categories:

- CPU binary translation. The hypervisor traps selected CPU instructions, such as CPU-privileged instructions[3] and performs necessary translations to make the guest OS think it has complete control over the server hardware.
- Hosted OS, application-layer abstraction. The software-layer abstraction resides as an application on top of a host OS. Most use binary translation and do not require a modified guest OS. This type of machine abstraction is typically used in smaller departmental environments rather than large data centers.
- Hardware-assisted virtualization (full virtualization). New processor hardware, such as AMD V or Intel VT x processor virtualization technologies that extend the x86 instruction set, assists the hypervisor. These new instructions move some supervisory control for the workload container into the CPU hardware layer.
- Paravirtualization. The guest OS is modified to include paravirtualized I/O drivers. The modified guest OS makes calls directly to the virtualized I/O services and other privileged operations, rather than having the hypervisor layer trap and translate all privileged instructions. As a result of the guest OS modification, the guest OS is now capable of being aware of its virtualized host (hypervisor).
- Hosted OS, kernel-layer abstraction. This method differs from the others because the abstraction technology is built directly into the OS kernel rather than having a separate hypervisor layer. As a result of the tight integration of the machine abstraction into the OS kernel and the assistance of AMD-V and Intel VT-x technologies, any host OS applications appear to run as peers to the virtual machine guest OS.

Appendices A and B provide additional descriptions of these techniques. Table 1 lists some of the most common vendor implementations and the primary hypervisor techniques used. Vendor

[2] Terminology varies when referring to software-based hypervisor technology. The definitions in this document may differ from other documents.
[3] Appendix A contains additional information about processor virtualization and privileged instruction calls.

implementations continue to evolve, and they may use more than one technique; therefore, this implementation list should be considered only as a sample of possibilities.

Table 1. Common vendor implementations of virtual machines

Virtualization technique	Examples of vendor implementations	Vendor URLs
CPU binary translation	VMware ESX	www.vmware.com
Hosted-OS, application-layer abstraction	Microsoft Virtual Server 2005 R2 VMware Server (formerly VMware GSX)	www.microsoft.com/virtualserver www.vmware.com/products/server
Hardware-assisted virtualization (full virtualization)	Citrix XenServer Windows Server 2008 Hyper-V	www.citrix.com/Xenserver http://www.microsoft.com/windowsserver2008/en/us/hyperv.aspx
Paravirtualization	Citrix XenServer Red Hat Enterprise Linux SUSE Linux Enterprise	www.citrix.com/Xenserver www.redhat.com/rhel/ www.novell.com/linux/
Hosted-OS, kernel-layer abstraction	Red Hat Kernel Virtual Machine	www.redhat.com/promo/qumranet/

Integrated hypervisors in ProLiant servers

ProLiant iVirtualization integrates select third-party hypervisors onto ProLiant platforms allowing customers to quickly set up and manage virtual machines. The first integrated versions available are VMware ESXi and Citrix XenServer. HP expects to offer an integrated version of the Microsoft Hyper-V Server hypervisor in late 2009.

The integrated hypervisor solution provides important technologies to administrators:

- Easily deploy and manage virtual machines using familiar ProLiant server management tools
- Move easily from stand-alone implementations to full enterprise versions of the hypervisor products
- Manage virtual machines on a single physical server using a simple graphical user interface, the ProLiant Virtual Console (currently available only with Citrix XenServer)

Use of familiar HP management tools

The integrated hypervisor solution models the virtual machine as an extension of the physical server, allowing ProLiant customers to easily deploy and manage a virtual machine in the same way they deploy and manage a physical server. ProLiant iVirtualization leverages customers' familiar ProLiant server management tools so that any of the hypervisor versions can be managed using HP Systems Insight Manager (HP SIM).

For example, in its ESXi hypervisor, VMware removed the SNMP agents and uses Web-based enterprise management WBEM providers instead.[4] When using the ProLiant-integrated VMware ESXi, the WBEM providers in HP SIM are pre-installed along with VMware ESXi. This method of distributing virtual machine technology removes time-consuming installation steps, simplifies host configuration,

[4] For more information about VMware ESXi, see the technology brief titled "HP VMware ESXi management environment," available at http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01570108/c01570108.pdf

and increases capacity expansion for customers who already use VMware's data center virtualization and management suite, VMware Infrastructure (VI). HP SIM 5.2 and later incorporates an improved installation process and supports association and discovery for VMware ESXi. Enhanced hardware alerting and inventory management allows customers to actively receive information about the health of their ProLiant servers.

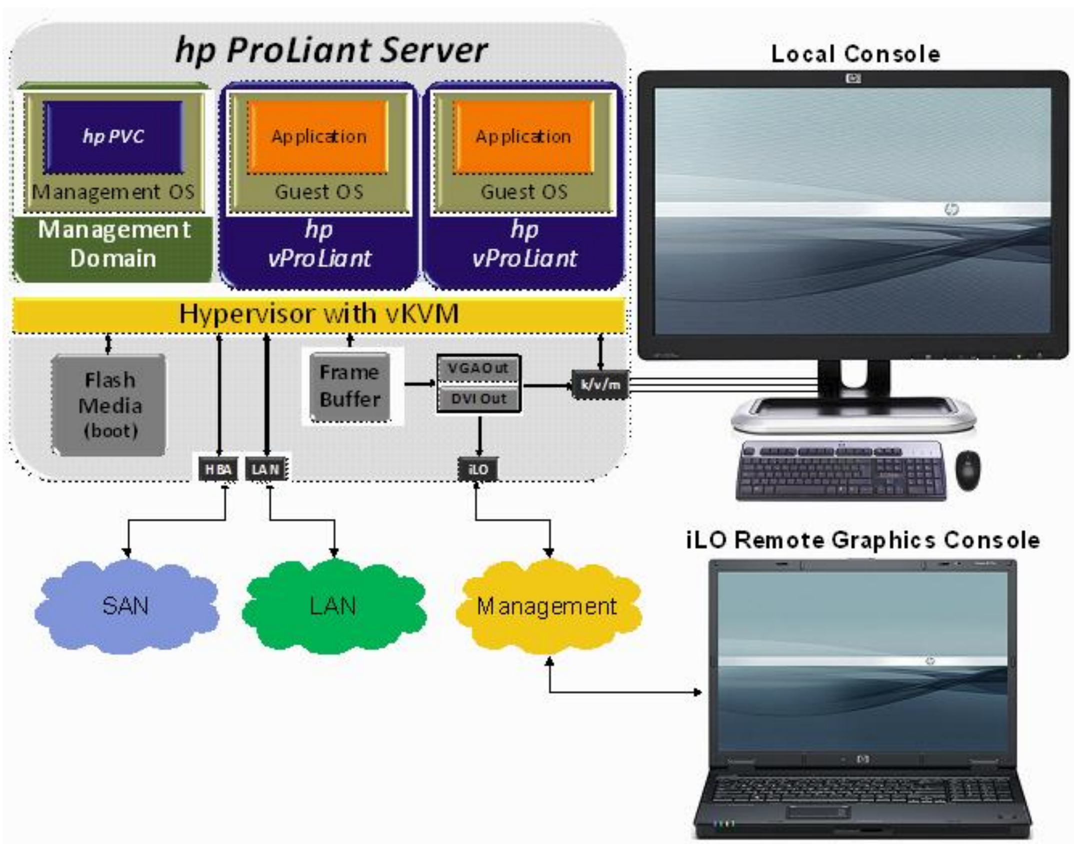
Scale-up to enterprise versions

The HP integrated hypervisors allow customers to grow their solutions from small implementations to the full enterprise versions. For example, because the drivers are included on the integrated solution, administrators can move from an installation of VMware ESXi to the complete VMware Infrastructure suite without removing and re-installing necessary drivers. HP partnerships with virtual machine solutions suppliers give administrators confidence that they will be deploying their virtual machines on the industry's broadest portfolio of servers, storage, software, and services designed to support machine abstraction.

ProLiant Virtual Console

HP has developed a ProLiant Virtual Console (PVC) that provides unique HP management capabilities. Designed for new users of virtual machine technology, HP PVC provides a simple graphical user interface for managing single-server virtual machines without the need for a separate management server (CMS) or even a network. PVC allows administrators to create and manage local virtual machines using the server's keyboard/video/mouse or by using HP iLO (Figure 3).

Figure 3. The Citrix XenServer integrated hypervisor includes the ProLIant Virtual Console for management.



The first implementation of the PVC is available in the integrated Citrix XenServer solution. HP is working with other partners to make use of this technology across the ProLiant portfolio as applicable.

For more information about integrated hypervisors, see the HP website www.hp.com/go/proliantvirtualization and see the technology brief titled "Integrated hypervisor virtualization technologies in HP BladeSystem ProLiant server blades and ProLiant servers" at http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01518167/c01518167.pdf.

Physical-layer abstraction: HP Virtual Connect solution

Virtual Connect is a physical-layer machine abstraction technology that parallels software-layer abstraction by allowing similar server workload flexibility and mobility. Just as hypervisor software abstracts physical servers into virtual machines, HP Virtual Connect technology abstracts groups of physical servers within a VC domain into an anonymous physical machine abstracted as PM-logical servers (PM-LS). With firmware version 1.31 and later, the embedded Virtual Connect Manager abstracts the server connection and network node ID/address (MAC and WWN) also in addition to defining the container and settings as part of a Virtual Connect server profile. Like VM-LS, PM-LS can be adjusted based on OS workload entitlements, while preserving Ethernet and SAN network connections of the OS boot image.

The overall Virtual Connect solution also uses the Virtual Connect Enterprise Manager (VCEM) software application, which provides central management of up to 150 Virtual Connect domains. VCEM aggregates Virtual Connect LAN and SAN addresses (node IDs) into a central repository and enables group-based administration of Virtual Connect domains, allowing customers to physically or logically link multiple enclosures, and to rapidly pre-provision, migrate, and recover server-to-network connections across the datacenter.

Virtual Connect consists of interconnect modules for Ethernet and Fibre Channel and includes embedded software for configuring single HP BladeSystem c-Class enclosures (VC Manager). The software and hardware present in the Virtual Connect modules provides the abstraction functions so the external network "sees" a pool of network connections within each BladeSystem enclosure.

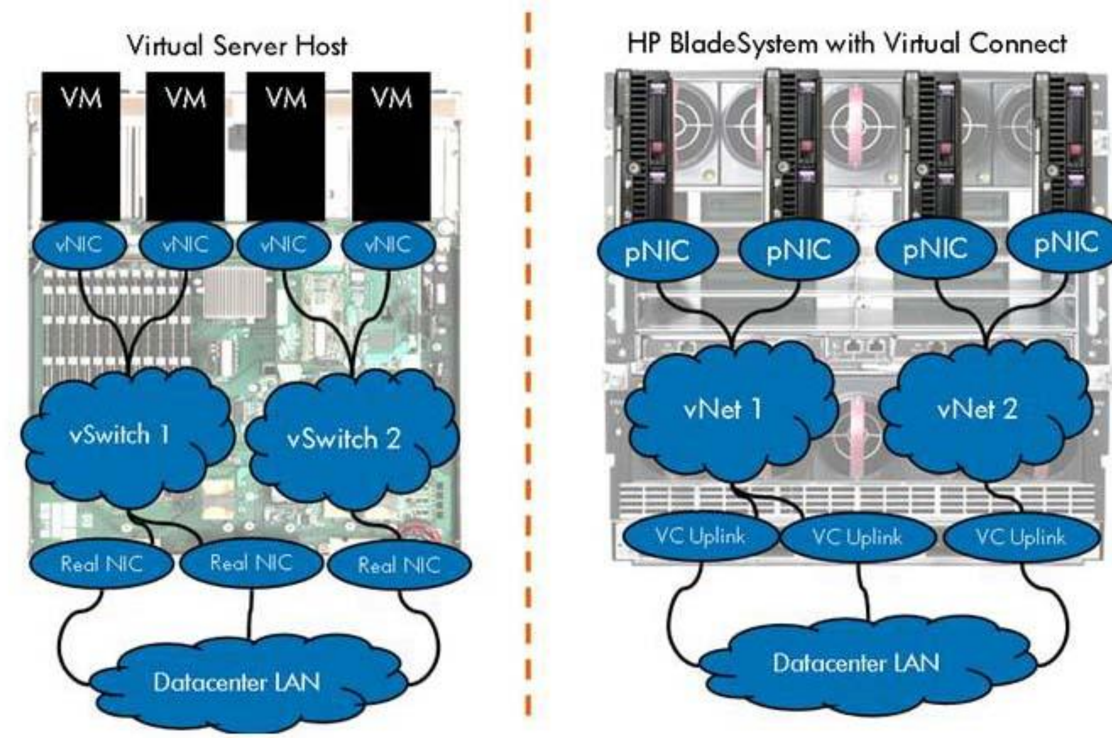
The most recent technology introduced as part of Virtual Connect is Flex-10 technology, which enables customers to partition a 10 gigabit (Gb) Ethernet connection and to regulate the size and data speed of each partition.

Virtual Connect server profiles

Instead of using the MAC addresses and WWNs created when the NICs or host bus adapters (HBAs) are manufactured, the Virtual Connect Manager creates specific server connection profiles, assigns unique MAC addresses and WWNs to these profiles, and then associates them with the BladeSystem enclosure bays so that these addresses are held constant. The server identity, or server profile, that Virtual Connect manages includes the server's LAN and SAN connectivity settings, Virtual Connect network assignments, managed MAC addresses and WWNs, PXE configuration and Fibre Channel boot parameters. In addition, Virtual Connect can define the abstracted container including the container logical IDs (logical serial number and logical Universally Unique Identifier, or UUID) and settings as part of the VC server profile.

This is conceptually the same process that a hypervisor follows when configuring virtual machines. In a virtual machine environment, the hypervisor defines the VM settings, logical ID including the network node address or virtual NIC identities for each virtual machine, and uses a software implementation of a layer 2 bridging device (a vSwitch) to provide external network connectivity to the virtual machines (left-hand side of Figure 4). With Virtual Connect, the Virtual Connect Manager defines the NIC identities for each physical server blade and uses a hardware implementation of a layer 2 bridge, called a Virtual Connect Ethernet network (vNet), to provide external network connectivity to the server blades.

Figure 4. Virtual Connect abstracts the networking connections in a manner similar to networking in a hypervisor environment.



For a more in-depth description of the similarities between the network abstractions of hypervisors and those of Virtual Connect, see the paper titled "HP Virtual Connect: Common Myths, misperceptions, and objections" available at http://bizsupport.austin.hp.com/bc/docs/support/SupportManual/c01598869/c01598869.pdf.

With Virtual Connect firmware releases of v1.31and later, customers can define the logical server abstractions within a VC server profile definition. These logical server abstractions go beyond the physical server connection and node ID to include the container definition, itself, with dynamic workload adjustments using Flex-10 technology. As a result, the VC server profile can reflect new levels of mobility and flexibility of the system image to the OS.

With the server identity being held constant through Virtual Connect, the following results occur:

- The underlying server networking hardware is transparent to the application/OS workload, allowing hardware components to be replaced or the OS workload to be moved to a completely different server blade. This improves flexibility for a physical or a virtual machine infrastructure.
- Network resources are provisioned in bulk, whether connected to physical machines or virtual machines.
- Virtual Connect separates the server infrastructure from the network and storage infrastructure, enabling a server administrator to add, move, or change servers without disrupting the external LAN and SAN networks.

For more information about Virtual Connect technology, see the technology brief titled "HP Virtual Connect technology implementation for the HP BladeSystem c-Class": http://h20000.www2.hp.com/bc/docs/support/SupportManual/c00814156/c00814156.pdf.

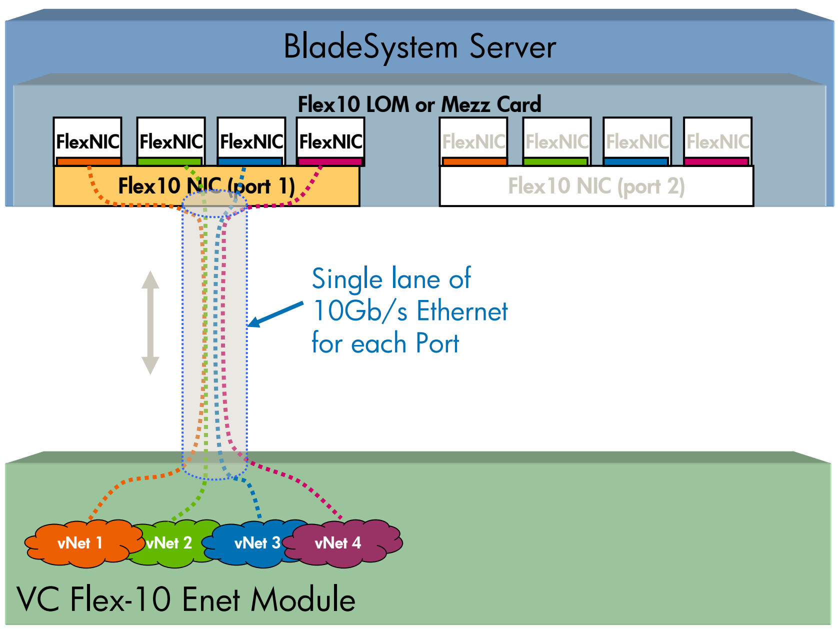
Flex-10 for Virtual Connect

Flex-10 technology is the newest extension of HP Virtual Connect technology. It is a hardware-based, I/O abstraction solution that allows customers to partition a 10 gigabit (Gb) Ethernet connection, to regulate the size and data speed of each partition, and to prioritize each of those partitions. Customers can configure a single 10Gb network port to represent up to four physical NIC devices, or *FlexNICs*, for a total bandwidth of 10Gbps.

Flex-10 hardware consists of two components: a 10Gb Flex-10 LAN-on-motherboard (LOM) device or 10 Gb Flex-10 mezzanine card; and the HP Virtual Connect Flex-10 10Gb Ethernet Module. The 10Gb Flex-10 LOM or mezzanine card connects to a 10Gb server port on select BladeSystem servers.

To the server OS (bare-metal OS), each of the four FlexNICs appears as a discrete NIC, with its own driver. The administrator defines the bandwidth available to each FlexNIC—from 100Mb to 10Gb—through the Virtual Connect interface. While the FlexNICs share the same physical port, traffic flow for each FlexNIC is isolated with its own MAC address and virtual local area network (VLAN) tags (see Figure 5). The Virtual Connect Flex-10 Ethernet Module recognizes Flex-10 connections from the server as part of a Virtual Connect server profile.

Figure 5. FlexNics share a physical link.



Flex-10 technology complements virtual machine technology by providing a large I/O bandwidth, multiple hardware-virtualized NICs, and the ability to change the bandwidth allocated to each FlexNIC.

Flex-10 is currently available with supported HP BladeSystem servers (for supported servers, see http://h18004.www1.hp.com/products/blades/components/ethernet/10-10gb-f/questionsanswers.html#c1). For more information about Flex-10 technology, see the technology brief "HP Flex-10 technology" at http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01608922/c01608922.pdf.

Virtual Connect Enterprise Manager

Built on the Virtual Connect architecture integrated into every BladeSystem enclosure, VCEM is a separate application that provides a central console on a CMS to perform efficient administration of LAN and SAN connections. VCEM provides group-based management that enables administrators to rapidly configure, assign, move, and failover server-to-network connections and their workloads for up to 150 BladeSystem enclosures (2400 servers). Administrators can logically link separate domains and move server profiles between Virtual Connect domains, as long as the servers are physically connected to the same networks and part of the same Virtual Connect Domain Group.

Together, Virtual Connect and Virtual Connect Enterprise Manager create a change-ready infrastructure that allows system administrators to add, replace, and recover servers across the data center in minutes without impacting LAN and SAN availability. This flexible infrastructure also provides foundation capabilities for the logical server management and other capabilities delivered with HP Insight Dynamics-VSE.

For more information about Virtual Connect Enterprise Manager refer to: www.hp.com/go/vcem.

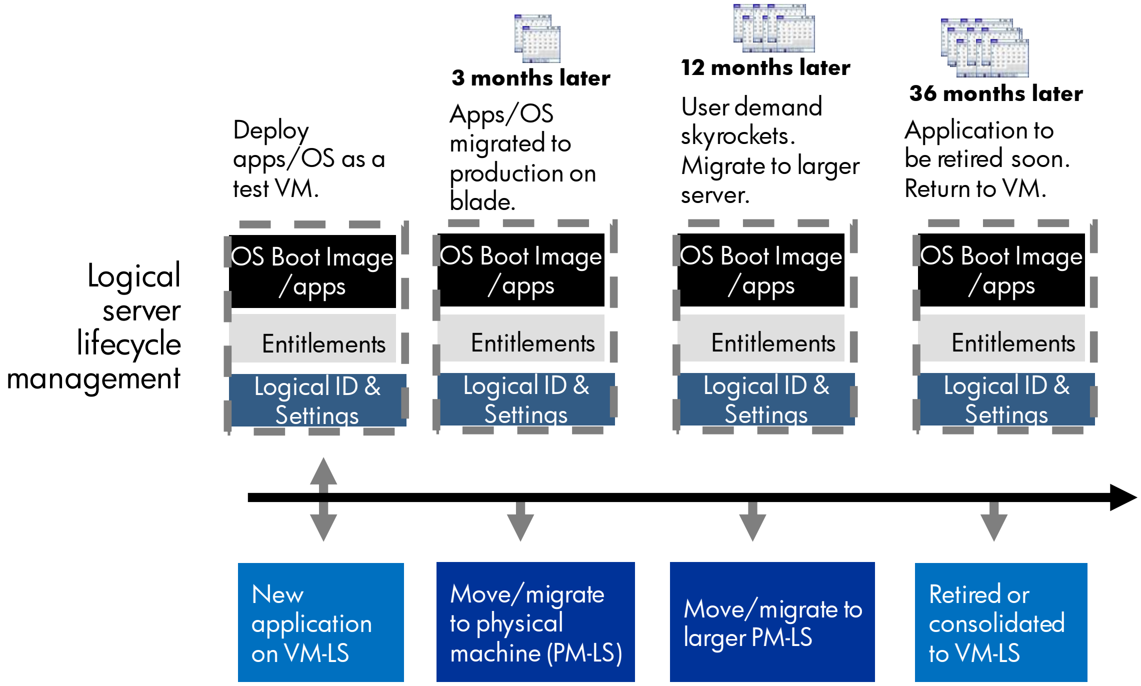
Logical servers

Logical servers are a new class of abstracted servers that allows administrators to manage physical machines and virtual machines using the same management construct. A logical server is defined by a server profile that can be created and migrated across physical and virtual machines. Migrations are limited to like-to-like as of this writing (V2V or P2P). Over time, migrations and moves will be made as seamless as possible.[5] A logical server profile includes the definition of the system services and resources—everything that the OS and application stack for a given workload require to operate—whether these services and resources are virtual, physical, shared, or unshared. The logical server profile can be managed centrally by means of a CMS using the Logical Server Manager viewer within the HP Insight Dynamics – VSE suite, (described in the following section). The profile can be applied to a virtual machine using software-layer abstraction (hypervisor technology as the VM host) or to a physical machine using physical-layer abstraction (HP Virtual Connect technology as the PM host).

The logical server approach lets administrators plan, deploy, migrate, adjust, and manage server instances in their environments regardless of how the logical server is created. The logical server allows flexibility and adaptability within the application/OS workload, minimizing any impact to shared network and storage domains (Figure 6).

[5]Currently the implementation of logical server migrations within Insight Dynamics – VSE must be like-to-like P2P or V2V. P2V and V2P migrations are available by means of an escape from the ID-VSE console to Server Migration Pack.

Figure 6. Logical servers streamline lifecycle management by enabling flexibility of the application/OS workload.



For additional information about logical servers, see the technology brief "Introducing logical servers: Making data center infrastructures more adaptive": http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01402013/c01402013.pdf.

HP Virtualization Management Software

As an IT infrastructure becomes more virtualized and abstracted, management can become an increasingly complex task. This is especially true when multiple types of machine abstraction exist in the data center (such as software-based VM-Logical Servers, hardware-based PM-logical servers, and Virtual Connect server-to-network abstraction). In such an environment, a unifying CMS becomes increasingly important to provide more consistent management:

- Improved visibility and understanding of workload placement
- Optimized resource scheduling and workloads
- Controlled from a single integrated console and coordination among multiple CMS systems

As multiple CMS systems may need to coordinate workloads, integration and partnership become more important across the data center. HP is uniquely positioned in the industry with its partner relationships, broad portfolio of in-house technology, and professional services to bring together comprehensive management solutions including a variety of programmatic and graphical user interfaces. The relationships that HP has established, such as those with hypervisor vendors, will enable HP to recommend partner solutions that are complementary to the HP solutions.

HP Insight Dynamics – VSE software

HP Insight Dynamics – VSE software provides an integrated manager for all logical servers, regardless of whether they are created using physical or virtual machines. HP Insight Dynamics – VSE incorporates other HP tools that have been used successfully for years to manage physical and virtual machines:

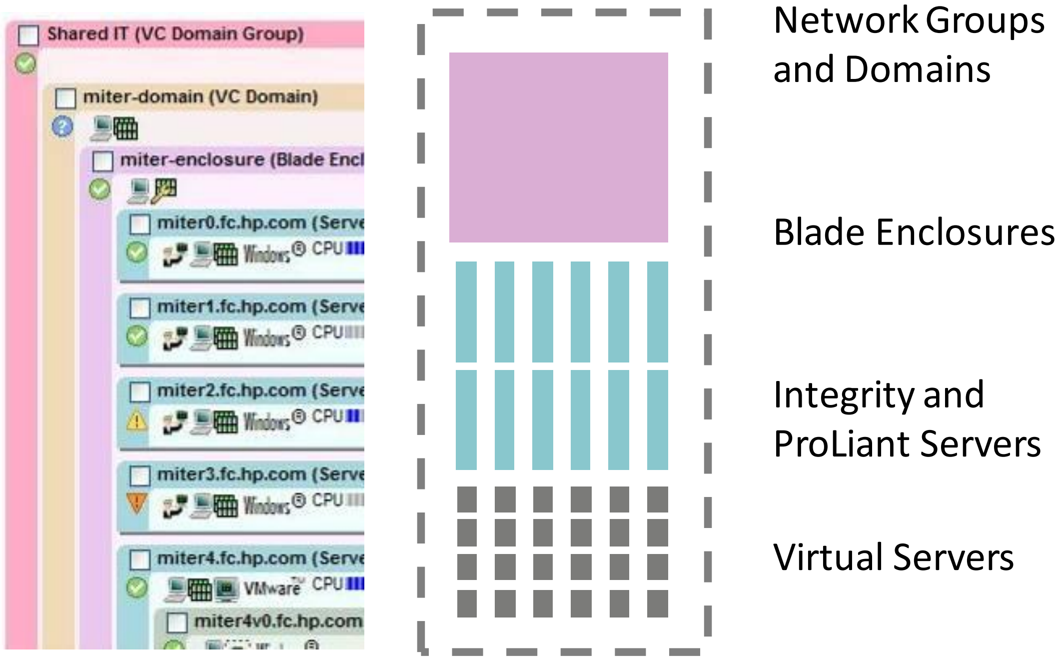
- HP Virtual Machine Management Pack, which provides central management and control for VMware, Microsoft, and Citrix virtual machines. It integrates with vendors' virtualization management tools, such as VMware vCenter Server (formerly known as VirtualCenter), and allows administrators to manage virtualization platforms from the same console used for physical server administration. (www.hp.com/go/VMM)
- HP Server Migration Pack – Universal Edition, which migrates existing servers and their content to the latest HP BladeSystem or ProLiant server technologies or the latest virtual machines from VMware, Microsoft, and Citrix. Server Migration Pack – Universal Edition facilitates migrations from physical to virtual platforms (P2V), from virtual to physical (V2P), or from one virtual machine to another (V2V). Essentially it lands all the drivers that the administrator would normally move manually, and then pulls all the existing data, applications, and operating system on top. (www.hp.com/go/migrate)

In addition, HP Insight Dynamics – VSE leverages the separate Virtual Connect Enterprise Manager (VCEM) application, which enables administrators to centrally deploy and manage multiple BladeSystem c-Class enclosures configured with Virtual Connect (Virtual Connect Domain). VCEM also provides a central pool of Virtual Connect LAN and SAN addresses (MAC and World Wide Names) and group-based management tools that allow system administrators to logically link separate domains and migrate VC server profiles between Virtual Connect domains that are part of the same Virtual Connect domain group. More information about VCEM is available at www.hp.com/go/vcem.

The logical server management interface provides a graphical view of the system topology, allowing administrators to easily visualize and manage all servers, their connectivity, and their unique server attributes. Administrators can use this tool to define and deploy logical servers, monitor logical servers, edit logical server profiles, define the access and control capabilities through role-based control, analyze capacities, and adjust logical servers to meet their needs.

Insight Dynamics – VSE includes integrated capacity planning technology. Administrators can view historical use data and pre-test workloads onto different sets of server resources. For example, they can carry out "what-if" scenarios, and evaluate performance, power implications, and consolidation scenarios. An administrator can drag and drop a logical server with immediate *best-fit* information through a five-star rating system—not merely drag and drop into the first virtual machine or physical machine that would fit a logical server profile. Figure 7 shows an example of the Insight Dynamics – VSE logical server management user interface and how it incorporates all logical servers, whether they are landed on physical machines or virtual machines.

Figure 7. HP Insight Dynamics – VSE user interface facilitates managing logical servers by controlling physical and virtual machines in the same way.



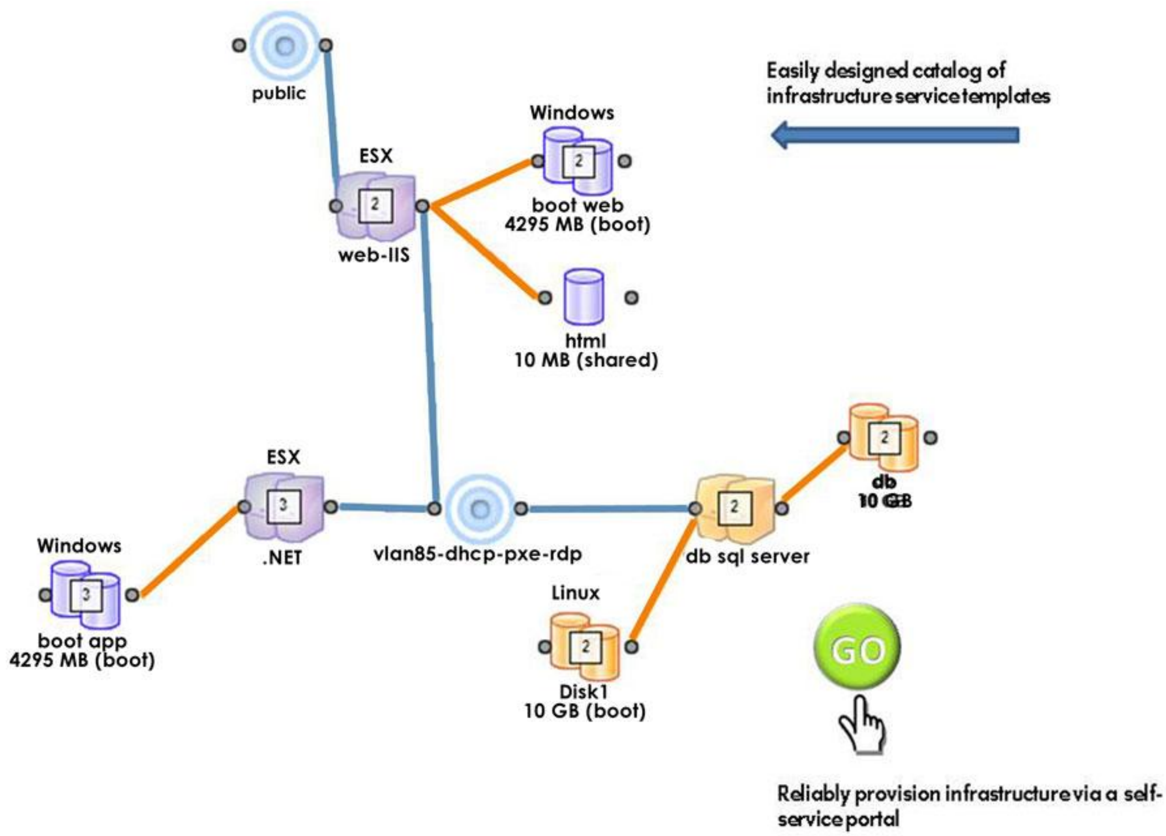
For more information about Insight Dynamics – VSE, refer to www.hp.com/go/insightdynamics.

HP Insight Orchestration software

HP Insight Orchestration extends the management capabilities of the HP Insight Dynamics – VSE suite for ProLiant servers to automate the provisioning and repurposing of an IT infrastructure. HP Insight Orchestration software enables administrators to rapidly provision a server infrastructure consistently and automatically from pools of shared resources using a self-service portal. Administrators can assign resources as varied as a single virtual machine to complex multi-tier environments with physical and virtual server and storage resources. The key capabilities of HP Insight Orchestration include the following attributes:

- Ability to visually design standardized infrastructures
- Ability to automatically provision using standardized templates, streamlined approval processes, and a self-service web portal. (Figure 8)
- Ability to seamlessly integrate with existing IT management tools and processes across the IT lifecycle—from design through operation and the ultimate repurposing of assets. It uses an embedded workflow automation tool, HP Operations Orchestration software, to make infrastructure delivery processes more efficient and reliable across IT functions, including the roles of IT architects, administrators, and operations teams.

Figure 8. The user interface for HP Insight Orchestration software includes a catalog of templates in a self-service portal.



For more information about Insight Orchestration, refer to www.hp.com/go/insightdynamics.

Conclusion

Virtual machine technology for x86-based servers is being incorporated into data centers at levels that are causing a true paradigm shift for IT administrators. HP is integrating third-party virtual machine technology into ProLiant platforms to offer out-of-the-box machine abstraction offerings from several different vendors. These integrated solutions include core HP server management tools that help customers optimize IT infrastructure, save time, and apply upgrades easily and reliably.

Because of their flexibility, ProLiant servers, and especially BladeSystem servers, provide a natural platform for virtual machine implementations:

- Hardware redundancy and availability
- Embedded intelligence with Integrated Lights-Out management and Onboard Administrator (BladeSystem)
- Capabilities for large memory, processing, and I/O footprints
- Wide range of storage options including boot-from SAN, shared storage, direct-attached hot-plug SAS drives, and Smart Array Controller options
- Power management tools:
  - Power meter for monitoring server power consumption use
  - Power Regulator for higher server efficiency

– High-efficiency power supplies
– Power capping feature for provisioning power to groups of ProLiant servers
– HP Thermal Logic (for monitoring and managing BladeSystem servers)

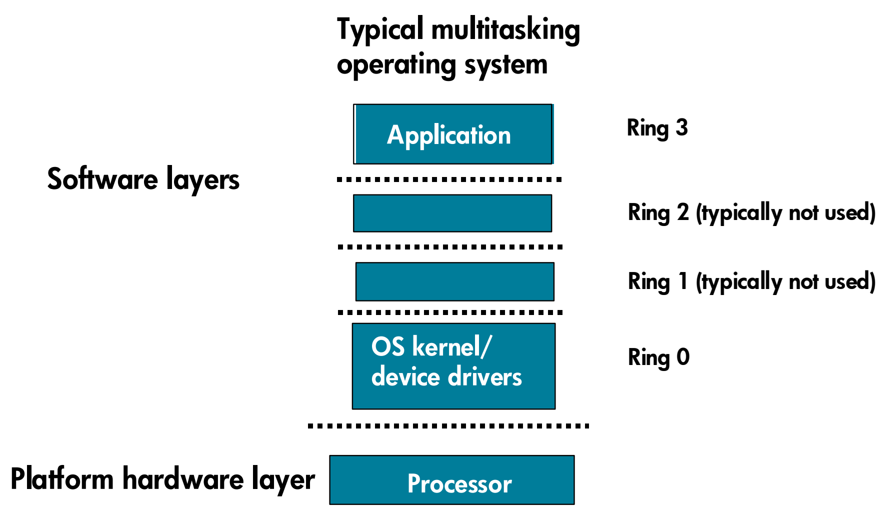
Furthermore, HP Virtual Connect technology in the BladeSystem c-Class architecture provides physical-layer abstractions, enabling machine abstractions at the server/network edge. The combination of HP Virtual Connect and virtual machine technology provides IT administrators with new capabilities for simplifying day-to-day operations, changing server resources quickly, and increasing productivity. With the addition of Flex-10 technology, administrators have a way to increase the number of NICs without the additional processor overhead or latency associated with software-based virtualization. Significant infrastructure savings are also realized since additional server NICs and associated interconnect modules may not be needed.

Virtual Connect also enables the physical layer abstraction of logical servers, so that administrators can use the concept of a logical server to manage both virtual and physical machines in a resource pool. HP continues to introduce new tools, such as HP Insight Dynamics – VSE and HP Insight Orchestrator software, that help ease server management, especially in heterogeneous environments of physical and virtual machines. HP plans to further expand the capabilities of server hardware to abstract other hardware resources inside the server. HP continues to move forward with machine abstraction technologies, working with internal development teams and with working groups from industry-standard bodies to develop new areas of abstraction and management for the next-generation Adaptive Infrastructure.

Appendix A: Background about processor ring layers

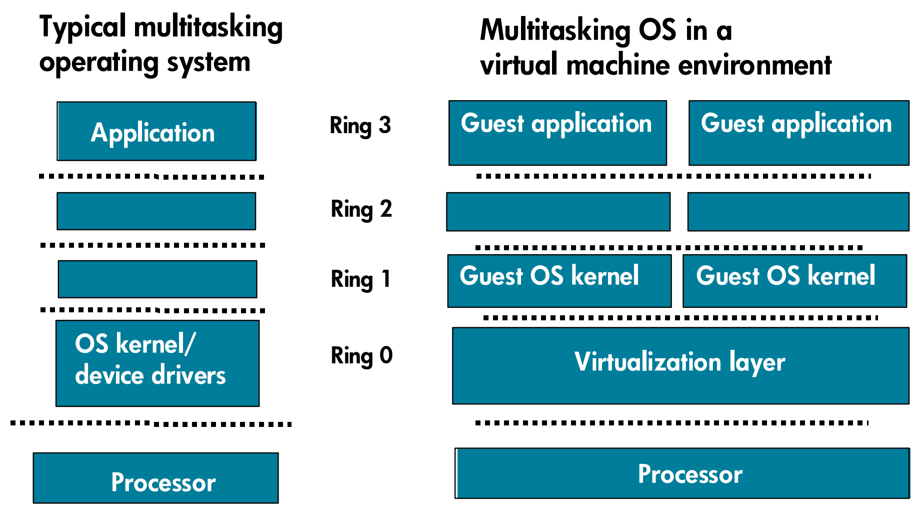
An x86 processor includes four different "ring" layers. These layers differentiate the type of instructions that the processor can execute, depending on what software is calling the instruction (Figure A1). In a typical multitasking OS, the OS operates in "ring 0," in which it has full access to all the processor and platform resources, such as memory mapping. This is the most privileged level, also referred to as kernel mode. Applications typically operate in "ring 3," sometimes referred to as user mode, in which functions such as memory mapping are restricted to keep one application from adversely affecting another. Originally, ring layers 1 and 2 were designed to house elements such as device drivers and the OS file system; however, these layers are not typically used today. The device drivers now typically reside with the OS in ring 0, and the file system resides with the application in ring 3 or with the OS in ring 0. The protected ring layers in the processor hardware work hand-in-hand with the OS to enable a multitasking OS to preempt an application off the processor, providing the appearance that multiple applications are operating simultaneously.

Figure A1. In a typical multitasking OS, the application resides in ring 3, while the OS kernel and device drivers reside in ring 0.



Because the hypervisor must be able to control all of the system's physical resources, just as the OS normally does (with privileged instructions), the hypervisor normally operates in ring 0 and moves the OS into ring 1 or ring 3 (Figure A2). However, this can create a conflict and potential excessive system faults when the OS issues a privileged instruction but it is not operating at the privileged ring 0 level. The result is that the hypervisor must act as a fault handler, executing many lines of code to determine what the OS has requested and handling the request as an intermediary between the OS and the hardware.

Figure A2. In a virtual machine environment such as VMware ESX Server that uses binary translation, the guest OS operates in ring 1 or ring 3.



With processor-assisted virtualization, hypervisor software will be able to run in a more privileged ring layer. This will allow guest operating systems to run in their normal ring 0 layer. In addition, AMD-V and Intel VT-x include new instructions that are specific to hypervisors. The combination of these two factors (higher-privilege ring level and new hardware instructions) may provide a more efficient way for hypervisors to virtualize server hardware.

Appendix B: Common types of software-layer abstraction for virtual machines

Binary translation

With binary translation technology, the guest OS is not aware it is operating on virtualized hardware. The hypervisor manages the access of each guest OS to the physical hardware resources by masking the hardware from the guest OS. It emulates portions of the system hardware and provides the guest OS with the illusion of a standard physical server with well-defined hardware devices. The hypervisor ensures that any instructions from the guest OS that affect system parameters—such as privileged instructions to the CPU—are handled in a way that does not affect the operation of other guest operating systems or cause OS kernel faults. The hypervisor traps the instruction and performs necessary translations that make the guest OS think it has complete control over the server hardware.

Hosted OS, application-layer abstraction

Another software-only approach uses a hypervisor layer that is hosted by an underlying OS. Because it resides as an application on top of the host OS, this type of abstraction inherits its hardware support and device compatibility from the host OS. This provides an advantage for customers who want to run an older, legacy OS on newer server hardware. However, the tradeoff for this hardware compatibility is the performance overhead required by the hypervisor layer. Typically, such hosted solutions are used in smaller, departmental environments rather than in large data center deployments because the hosted solutions often lack capabilities such as dynamic load balancing or clustering.

Hardware- assisted virtualization (full virtualization)

With hardware-assisted virtualization (sometimes referred to as full virtualization), the hypervisor is assisted by the processor hardware such as AMD-V or Intel VT-x processor virtualization technologies. In this scenario, when the guest OS makes a privileged instruction call, the processor (CPU) traps the instruction and returns it to the hypervisor to be emulated. Once the operation is serviced by means of the hypervisor, the modified instruction is returned back to the CPU for continued execution. Hardware assistance reduces the software overhead required by the hypervisor.

Hardware assistance from AMD-V and Intel VT-x technologies extends the x86 instruction set with new instructions that affect the processor, memory, and local I/O address translations. The new instructions enable guest operating systems to run in the standard Ring-0 architectural layer, as they were designed to do, removing the need for ring compression.

Paravirtualization

Paravirtualization refers to a technique in which the guest OS includes modified (paravirtualized) I/O drivers for the hardware. Unlike a binary translation approach, the hypervisor does not need to trap and translate all privileged layer instructions between the guest OS and the actual server hardware. Instead, the modified guest OS makes calls directly to the virtualized I/O services and other privileged operations. Therefore, paravirtualization techniques have the potential to exhibit faster raw I/O performance than binary translation techniques.

Some of the hypervisor implementations that use this method (Citrix XenServer, Red Hat Enterprise Linux 5, and SUSE Linux Enterprise) are unique in that they support paravirtualization when using a modified guest OS and hardware-assisted virtualization when the guest OS is not virtualization-aware.

Hosted OS, kernel-layer abstraction

Kernel-layer abstraction refers to a technique in which the abstraction technology is built directly into the OS kernel rather than having a separate hypervisor layer. This approach can still follow the requirements of hardware-assisted (AMD-V or Intel VT-x virtualization technologies) and paravirtualization, but removes the discrete hypervisor layer. The direct access to hardware could potentially provide greater performance than using a binary translation technology; however, because there is no separation between the hypervisor the operating system, there is the possibility that resource conflicts may occur between multiple virtual machines.[6]

[6] See the article titled "Xen proponents question merits of Red Hat KVM hypervisor" at http://searchservervirtualization.techtarget.com/news/article/0,,sid94_gci1318772,00.html; and for the opposing view, see Qumranet's white paper titled "KVM - Kernel-based Virtualization Machine" at www.qumranet.com/files/white_papers/KVM_Whitepaper.pdf.

For more information

For additional information, refer to the resources listed below.

Source	Hyperlink
HP webpages:	
BladeSystem servers	www.hp.com/go/bladesystem
BladeSystem Virtual Connect	http://www.hp.com/go/bladesystem/virtualconnect
Virtual Connect Enterprise Manager	www.hp.com/go/vcem
ProLiant servers	www.hp.com/go/proliant
HP Virtualization solutions	www.hp.com/go/virtualization
Industry-Standard Server Technology Communications	www.hp.com/servers/technology/
Insight Control Management Software	www.hp.com/servers/proliantessentials
Insight Dynamics - VSE	www.hp.com/go/insightdynamics
ProLiant iVirtualization	www.hp.com/go/proliantvirtualization
VMware and HP	www.hp.com/go/vmware
Citrix virtualization	http://www.citrix.com
Microsoft Virtualization Solutions	www.microsoft.com/virtualization/default.mspx
Novell (SUSE Linux Enterprise 10)	www.novell.com/linux/
Processor virtualization information	
AMD website	http://enterprise.amd.com/us-en/AMD-Business/Business-Solutions/Consolidation/Virtualization.aspx
Intel website	www.intel.com/technology/virtualization/index.htm
Red Hat Enterprise Linux 5	www.redhat.com/rhel/

Call to action

Send comments about this paper to TechCom@HP.com.

© 2006, 2008, 2009 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

AMD is a trademark of Advanced Micro Devices, Inc.

Intel is a registered trademark of Intel Corporation in the U.S. and other countries.

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

TC090107TB, January 2009