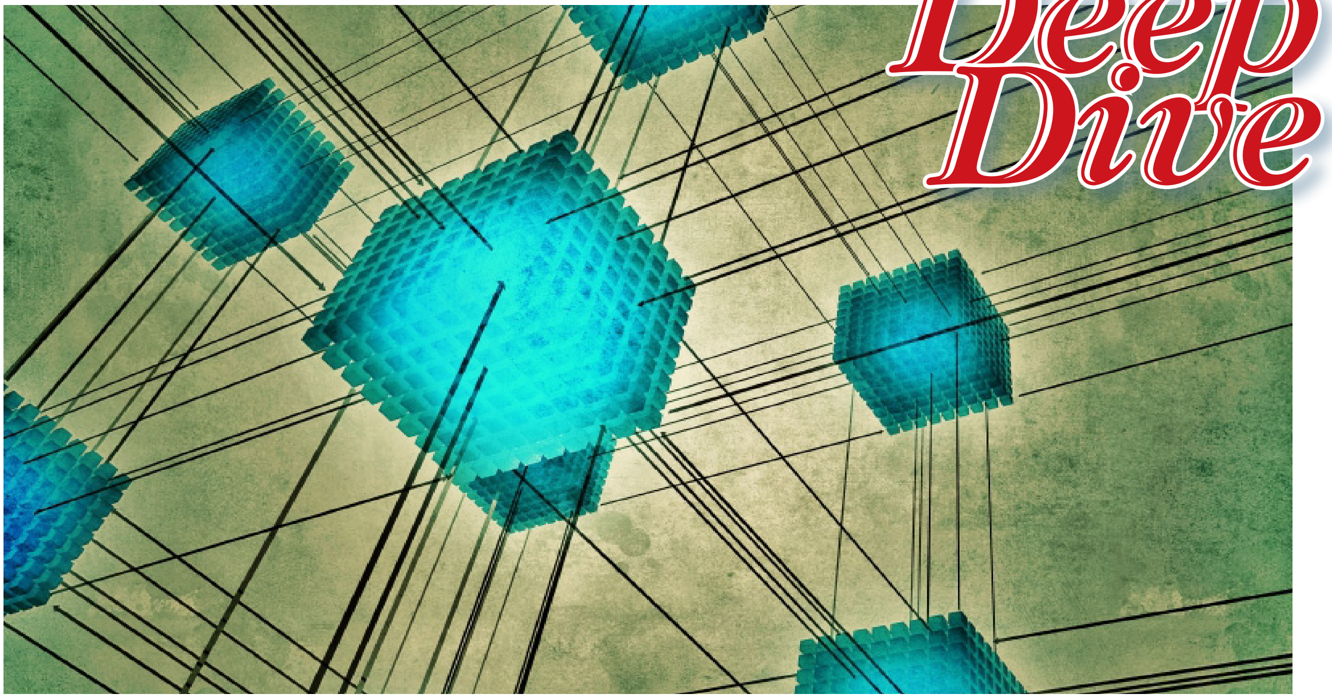
InfoWorld

SPECIAL REPORT

Networking for Virtualization

Deep Dive



Optimize your infrastructure for server virtualization

Preparing for today's virtualization demands

As virtualization has matured, the demands on network infrastructure have risen. Here's how to configure your network for ever-increasing server density

By Paul Venezia

ON THE SURFACE, connecting virtual servers to the network is much the same as it's always been: You configure appropriate IP addressing and routing information to a server network interface and the server can communicate. But the foundations of how that communication actually occurs have changed in the virtualized world.

Modern virtualization hosts can handle many more virtual servers than ever before -- and they need to be able to support the network demands for all of those instances. In both performance and configuration, this modern landscape requires a new depth of understanding, particularly in configuring virtualization hosts to maximize the reliability, bandwidth, and ease of management of the entire infrastructure.

As virtualization technology has progressed, the fundamental design of the network is not as straightforward as it used to be. New network virtualization technologies in today's hypervisors are turning traditional networking on its head -- and some of the more advanced capabilities of modern hypervisors can be obtained only by moving to non-traditional network designs.

As with just about everything in IT, the network is the key to the castle. Understanding and designing a stable, capable, and robust network ready for modern virtualization will avoid many potential bumps in the road.

BUILDING THE NETWORK CORE

LANs haven't changed fundamentally in many years -- except for dramatic declines in the cost per port and the ever-increasing bandwidth available on both copper and fiber links. These trends have made it possible to obtain the robust networking hardware necessary for virtualization networking without breaking the bank.

As you consolidate servers into a virtual infrastructure, you necessarily place many critical systems on a small number of physical hosts, so providing those hosts with as much redundancy, bandwidth, and connectivity options as possible will only serve to enhance the capabilities and reliability of the entire endeavor.

In a virtualization scenario of any reasonable size, multiple core switches with redundant inter-switch links are highly recommended, with (optimally) enough ports to support fully redundant virtualization host links on every planned network. Further, these switches should support Layer-3 networking, HSRP (Hot Swap Routing Protocol) or VRRP (Virtual Routing Redundancy Protocol), link aggregation, VLANs and VLAN trunking to fully realize the benefits of server virtualization. Together these features will help sustain performance, resiliency, and ease of management. Even in software-defined networks that use technologies such as VMware's VXLAN (Virtual eXtensible LAN), these foundational elements are critical.

A common scenario for a midsize virtualization platform might be a few modular switches with either a pile of gigabit copper ports or a smaller number of 10G copper or fiber Ethernet ports. These switches should be interlinked with redundant connections to protect against link failure, and be able to support the aforementioned link aggregation and trunking.

If the infrastructure is large enough, additional switches may accommodate racks of servers, and those switches should again be redundantly linked to the core and support the expected bandwidth requirements of the entire rack. The topology of this network is relatively simple, with each edge switch having multiple redundant connections to the core, using a spanning-tree layout to sort out the network paths and provide for failover switching.

BASIC BANDWIDTH DECISIONS

One of the first major choices you'll face involves the size of the pipe to connect the virtualization host servers to the network. Should you opt for 1G or more expensive 10G links? It all comes down to economics. If you have the budget to deliver multiple 10G links to each virtualization host, then by all means, go that route. If it's a stretch, you'll need to closely examine the realistic bandwidth requirements of the planned infrastructure in order to determine the suitable number of bonded 1G links for each server.

In many cases, the network and storage requirements of each host server may not demand more than two or four bonded 1G interfaces, which can reduce your expense considerably. On the other hand, if you implement 10G now, you can future-proof your investment and be ready for I/O growth in the near and distant future. It all depends on the cost of the 10G ports on both the switch and the physical server side.

Generally speaking, a virtualization host server will be adequately served by two 10G interfaces configured for redundancy, or six to eight 1G interfaces, depending on storage networking configurations. Remember that 1G Ethernet interfaces are cheap on both the server and the switch side, so don't skimp.

Another consideration in the 10G debate is whether to use 10GBase-T copper interfaces and cabling, or SFP+ connections. Currently, 10GBase-T switches tend to be more expensive than their SFP+ counterparts, but some of that savings is lost in the higher cost of SFP+ Twinax cabling generally used to connect servers, switches, and storage within a rack or groups of racks.

10GBase-T runs over standard Category 6 twisted-pair Ethernet cabling with standard RJ45 connectors on each end. However, for longer runs, Category 6a cabling should be used. This is more expensive than Category 5 or Category 5e cabling that has supported 100Mb and 1Gb networking for years. Make sure that any patch panels involved in the design are also Category 6/6a as well.

DEPLOYING VLANS

The use of VLANs in a virtualized infrastructure is a must unless you're moving toward virtualized networking such as VXLAN. By slicing up the network and trunking VLANs to the hosts, you can drop a VM on any network you wish, at any time. You can also use the enhanced control afforded by VLANs to constrain certain VMs to their own specific network, and in some cases, create airlocked VLANs that can be used in development environments. There's no reason in this day and age not to trunk to your virtualization hosts.

For instance, you may have a VLAN for front-end server traffic, a VLAN for back-end traffic, a management VLAN, and a storage VLAN if NFS or iSCSI storage is in use. You may also need one or more desktop VLANs for VDI (virtual desktop infrastructure) -- or any number of other scenarios. Slicing up a network into VLANs can simplify management while providing logical separation of network traffic, which can then benefit from simple QoS rules to ensure proper operation during periods of network congestion.

NETWORK VIRTUALIZATION

Beyond the traditional VLAN-driven network model is the relatively new concept of overlaying a traditional Layer-2 network with encapsulated network traffic to contain and control that traffic.

This method relies on the hypervisors rather than the network switches to handle the traffic separation and direction. Instead of configuring a VLAN on a switch or set of switches, you configure a distributed virtual switch across multiple hypervisors, and then overlay your different networks there.

This creates a network of multiple virtual networks that are all overlaid on a basic Layer-2 transport between hypervisors. The hypervisors accomplish this by encapsulating each packet before placing them on the physical network. This encapsulation instructs the network to deliver the packet to another hypervisor that then strips the encapsulation and delivers the original packet to the destination virtual server.

The upside of this is that you can configure and control nearly all aspects of the network from within the hypervisor, including firewalls, load balancers, virtual

LANs, gateways, and more. Additionally, multiple sites with multiple infrastructures can be stitched together to provide seamless transition of virtual servers between sites, without the need to change IP addresses or routing. To fully achieve a seamless site-to-site migration, for instance, you would need to implement VXLAN, or another, similar technology such as NVGRE (Network Virtualization using Generic Routing Encapsulation).

STORAGE NETWORKING FOR VIRTUALIZATION

A critical component of standard virtualization networking is storage networking. The three main methods of delivering shared storage to a virtualization host are: iSCSI, NFS, or Fibre Channel. iSCSI and NFS both use standard Ethernet to deliver storage resources, whereas Fibre Channel uses its own host-bus adapters and switches to completely isolate storage networking from the Ethernet network. The decision of which storage delivery technology to use generally hinges on both the current infrastructure in place and budget dollars available.

Fibre Channel storage will generally be faster and lower-latency than iSCSI or NFS, but the instances where these benefits turn out to be necessary in the real world are small in number. In most cases, the iSCSI or NFS services provided by a suitable storage array can accommodate all but the most heavily exercised virtualization infrastructures. That said, if a Fibre Channel storage infrastructure is already in place, it may be more reasonable to leverage that investment for a virtualization plan.

If you choose iSCSI, you'll need to make some decisions about the host hardware. Most virtualization solutions use software iSCSI initiators as the default, but some can make use of iSCSI accelerators that offload iSCSI header processing at the NIC level, leaving the CPU to handle the actual VM workloads. If at all possible, choose the iSCSI accelerators. Cycles saved on virtualization hosts equals cycles available to the virtual servers, which can directly lead to more VMs per host. And of course, iSCSI accelerators do a good job of enhancing iSCSI performance.

Note that an iSCSI deployment is highly dependent on the choice of switches to handle the iSCSI traffic. Some older 1G switches can have problems with iSCSI performance under load, whereas newer switches tend to be tuned for iSCSI traffic. Make sure that the switches in use for an iSCSI solution support jumbo frames and enable jumbo framing on the switches, storage array, and the host itself. The benefits derived from jumbo framing on iSCSI can be significant, as the larger packet sizes lead to reduced latency and better throughput.

If NFS is the choice, then there isn't much to do on the host side, as the storage array will be busy handling the NFS serving, with the virtualization host acting as the client. There are definitely benefits to using NFS as a VM store, such as ease of backup and restore. But there are detriments as well -- notably, the inability for some virtualization solutions to create raw disk mappings on NFS-based storage.

In some cases, you may find that NFS-based solutions are slightly faster and more responsive than iSCSI solutions, but that is largely based on the VM workload, the storage array itself, and the network in use. It's an excellent idea to conduct your own lab tests to determine which will work best for you. Smaller, less transaction-oriented infrastructures may be a better fit for NFS, and high-transaction environments may be better off with iSCSI.

BUILDING THE PHYSICAL HOST

When you build a virtualization host, keep in mind that you're providing a platform for many virtual servers. The first question when starting actual construction of the physical host is the choice of 1G or 10G Ethernet. If it's 10G, then in most cases it's as simple as connecting two 10G links to the core switches with VLAN trunking and link aggregation, or just link aggregation if you're using VXLAN or similar.

In the case of 1G links, it's best to divvy up the links across physical interface adapters. For instance, a basic virtualization host will likely have six 1G interfaces: two for front-end communication, two for back-end communication and VM migration purposes, and two for iSCSI, NFS, or Fibre Channel storage. Each of these pairs should be bonded using link aggregation and at least the front-end pair trunked to allow multiple VLANs. If the front-end I/O requirements are greater than a pair of 1G links, then the number should be increased, but the minimum should be two.

The back-end links are generally configured the same way. You need at least two for redundancy; more if you plan on heavy inter-VM communications. These links should be balanced with link aggregation as well, especially if they're destined to be communicating with multiple back-end hosts.

A word on link aggregation: Basic link aggregation, such as Cisco's EtherChannel, determines the link use based on MAC address. That is, once a pair of MAC addresses begin communicating, they will use only a single link in the bundle. New connections to different MAC addresses may use a different link, but if there's only a single pair of MAC addresses communicating, only one link will be used at a time. This means that you'll only get a single link per server interaction, potentially wasting the others if no other communication is occurring. When dealing with VM migration traffic, this means that when one host is migrating a single VM to another host, it will only utilize one link in the bundle. If that host is migrating multiple VMs at a time to different hosts, they will use multiple links. This is an important distinction, as it may result in one link in the bundle carrying the bulk of the traffic, with the others lightly utilized. However, by constructing the back-end network in this way, you also gain the benefit of failover; should one link go dark, another will pick up the traffic, so it's a better bet than a simple failover configuration.

This leaves the storage networking links. For iSCSI and NFS, these will be standard Ethernet, or portions of a 10G interface. If 1G iSCSI is the plan, then investigate the aforementioned iSCSI accelerators. Either way, these links should also be aggregated for redundancy and expanded bandwidth if multiple datastores are in use.

To sum up the basic host configuration: six 1G interfaces in bonded pairs for front-end, back-end, and storage links. For a 10G-equipped host, we'll have just the pair of 10G links configured in a failover or bonded pair.

In an ideal situation, both of these hosts would have an additional 1G interface configured for management traffic.

A WORD ON SECURITY DOMAINS

In many cases, a virtualized infrastructure will be running virtual servers that require connections to untrusted networks. For example, a mail relay or Web server may need to be placed on a DMZ network rather than the internal network, but will need to run on the same virtualization host as internal servers. This presents a bit of a quandary, with special consideration required for proper and secure implementation.

In a traditional switched network, there are three ways to approach this problem. Arguably, the more secure way is to dedicate one or more interfaces on each host to the untrusted network. This is obviously a bit more costly, because it requires additional interfaces on every host that may need to run these virtual servers, but it also explicitly pins the network interface of each VM to a physically separate interface on the host.

Another option is to trunk the untrusted network through the existing front-end interfaces on the host. This requires mixing physical security domains and bringing the DMZ network into the internal core switching rather than onto a dedicated DMZ switch. This can be cheaper to implement, but also requires substantial trust in the internal switching -- and in the network administrators, since a single miscue on the internal switch configuration can mix the untrusted and trusted networks within the switch, resulting in untrusted traffic mingling with trusted internal traffic. Either solution will work. The former is generally the best bet for security, although not for flexibility.

The third solution is simple, but the most expensive of all: A dedicated virtualization host farm just for DMZ or untrusted hosts. If a significant number of hosts require presence on untrusted networks, this may be a viable option, but for general purposes, it's overkill.

If you're using a virtual network overlay, then all of the switching, routing, and firewalling occurs within the hypervisors themselves. You would use the interfaces as you would for internal networks, but define those untrusted or lightly trusted networks within the hypervisor's network configuration.

ADVANCED CONCEPTS

The above is a baseline for virtualization networking. By building hosts this way, you're essentially guaranteed that the virtualized infrastructure has everything it needs for stable, reliable, high-performance networking. That said, some advanced concepts can fine-tune performance in infrastructures with heavier requirements.

In traditional networks, you can use virtual managed switching. Generally speaking, virtual server networking within the host is simple; each VM has a unique MAC address and is presented as a basic host to the switch on the other end of the link, as if the virtualization host server were an unmanaged switch. There is no realistic way in that scenario to control or constrain network parameters for each virtual server at the host level.

Virtual switches such as Cisco's Nexus 1000V can provide those features by functioning as a software-based switch running within the hypervisor itself. This essentially turns the unmanaged switch internal to the virtualization host into a managed switch, allowing virtual server switchports to be configured and managed like physical switchports are for physical servers. In addition, virtual switches can communicate with other instances in the infrastructure. When virtual servers migrate between hosts, their virtual switchports follow, allowing for specific switchport configurations to be maintained and monitoring to be consistent to the virtual server, not the host it runs on.

There are also a plethora of network virtual overlay technologies, such as the aforementioned VXLAN and NVGRE. Several virtualization and networking vendors have developed their own version of this concept, and there's no firm determination on which one may eventually become the industry standard. This means that if you choose to go that route, you will be using the technology best represented by your chosen hypervisor and network equipment.

Not every infrastructure will require the enhanced capabilities that virtual switches provide, but they can provide a significant manageability boost in some infrastructures.

QUALITY OF SERVICE

Although virtual managed switches can provide enhanced features, some advanced features may be built in to the virtualization solution itself. For instance, you may have the ability to implement QoS rules to each network link or bundle on each host. The depth and reliability of this option varies from vendor to vendor, but you may be able to implement guaranteed and burst rate limits on a per-VLAN basis, ensuring that servers communicating on critical networks are guaranteed to pass traffic during periods of high congestion. In many environments this isn't required, but it's worth investigating.

HARDWARE-BASED TRAFFIC MANAGEMENT

In some cases, your choice of hardware may add substantially to your networking capabilities. Some blade vendors, for instance, offer 10G virtual network interfaces that can be centrally configured within the blade chassis itself. For example, a pair of internal 10G interfaces to a blade can be configured within the blade chassis into four virtual interfaces that are then presented to the virtualization host as physical interfaces. This can be used to carve up bandwidth on that 10G interface into front-end, back-end, and storage networks with the chassis-based switch controlling the bandwidth allocations.

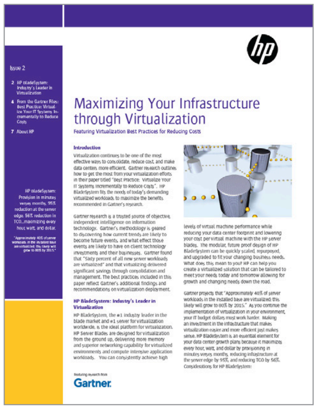
You could then dedicate, say, 4G of that 10G pipe to iSCSI traffic, 4G to back-end networking, and 2G to front-end networking. This takes the quality-of-service onus off the virtualization host and places it on the chassis switch, which can provide performance and management benefits.

CABLES EVERYWHERE

The basis of virtualization networking is that you'll find that you either need 10G networking or a large number of 1G interfaces on each virtualization host. Given that each host must have identical connections to each network, that can quickly add up to a large number of 1G switchports. With the price of 10G switching dropping, a newly planned virtualization environment will definitely benefit from the enhanced bandwidth and simplicity 10G provides.

That said, 1G still has a place in this game and will for some time to come. No matter how you ultimately decide to build your virtualization networking infrastructure, remember that although virtualization can greatly simplify enterprise computing, it can also cause a large number of headaches when physical hosts fail or have other problems that can suddenly affect a large number of virtual servers. If there ever was a perfect application of the old adage "a stitch in time saves nine," then virtualization is it. Go forth and plan accordingly.

Server virtualization resources



Issue 2

2 HP BladeSystem: Industry's Leader in Virtualization

4 From the Gartner Files: Best Practice: Virtualize Your IT Systems Incrementally to Reduce Costs

7 About HP

HP BladeSystem: Provision in minutes verses months, 95% reduction at the server edge, 56% reduction in TCO...maximizing every hour, watt, and dollar.

"Approximately 40% of server workloads in the installed base are virtualized; this likely will grow to 80% by 2015."

Maximizing Your Infrastructure through Virtualization

Featuring Virtualization Best Practices for Reducing Costs

Introduction

Virtualization continues to be one of the most effective ways to consolidate, reduce cost, and make data centers more efficient. Gartner research outlines how to get the most from your virtualization efforts in their paper titled "Best Practice: Virtualize Your IT Systems Incrementally to Reduce Costs". HP BladeSystem fits the needs of today's demanding virtualized workloads to maximize the benefits recommended in Gartner's research.

Gartner research is a trusted source of objective, independent intelligence on information technology. Gartner's methodology is geared to discovering how current trends are likely to become future events, and what effect those events are likely to have on client technology investments and their businesses. Gartner found that "Sixty percent of all new server workloads are virtualized" and that virtualizing delivered significant savings through consolidation and management. The best practices included in this paper reflect Gartner's additional findings and recommendations on virtualization deployment.

HP BladeSystem: Industry's Leader in Virtualization

HP BladeSystem, the #1 industry leader in the blade market and #1 server for virtualization worldwide, is the ideal platform for virtualization. HP Server Blades are designed for virtualization from the ground up, delivering more memory and superior networking capability for virtualized environments and compute intensive application workloads. You can consistently achieve high levels of virtual machine performance while reducing your data center footprint and lowering your cost per virtual machine with the HP server blades. The modular, future proof design of HP BladeSystem can be quickly scaled, repurposed, and upgraded to fit your changing business needs. What does this mean to you? HP can help you create a virtualized solution that can be tailored to meet your needs today and tomorrow allowing for growth and changing needs down the road.

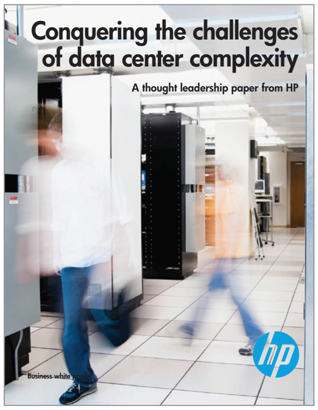
Gartner projects that "Approximately 40% of server workloads in the installed base are virtualized; this likely will grow to 80% by 2015." As you continue the implementation of virtualization in your environment, your IT budget dollars must work harder. Making an investment in the infrastructure that makes virtualization easier and more efficient just makes sense. HP BladeSystem is an essential element for your data center growth plans because it maximizes every hour, watt, and dollar by provisioning in minutes verses months, reducing infrastructure at the server edge by 95%, and reducing TCO by 56%. Considerations for HP BladeSystem:

Featuring research from Gartner.

HP NEWSLETTER WITH GARTNER RESEARCH:

Maximizing Your Infrastructure through Virtualization

Download now >>



Conquering the challenges of data center complexity

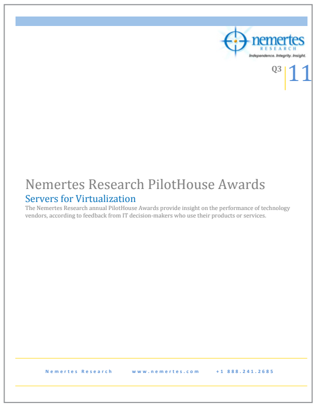
A thought leadership paper from HP

Business white paper

HP PAPER:

Conquering the challenges of data center complexity: A thought leadership paper from HP

Download now >>



nemertes RESEARCH Independence. Integrity. Insight.

Q3 | 11

Nemertes Research PilotHouse Awards

Servers for Virtualization

The Nemertes Research annual PilotHouse Awards provide insight on the performance of technology vendors, according to feedback from IT decision-makers who use their products or services.

Nemertes Research www.nemertes.com +1 888.241.2685

NEMERTES RESEARCH PILOTHOUSE AWARDS:

Servers for Virtualization

Download now >>

Sponsored by