How to Virtualize

Executive Summary

Data center virtualization projects have a ripple effect throughout an organization's infrastructure and operations. Collapsing multiple physical devices into software affects the people and processes supporting the data center. Before embarking on data center virtualization deployments, enterprises should account for how business processes, administrative rights, capacity planning, performance monitoring tools and security strategies will need to change.

at&t

Introduction

Virtualization projects are under consideration or already in progress at many organizations looking to decrease their capital and real estate expenses and to launch energy-conservation initiatives. Through the use of special virtualization products and services, data center managers can apportion computing hardware resources to operating systems and applications on a time-sharing basis. In doing so, they can significantly boost the utilization and efficiency of servers, network equipment and storage devices. Such projects reduce the overall number of physical devices needed, the floor and rack space needed to house them and physical equipment management requirements.

Virtualization, then, holds a number of potential capital and operational cost-saving benefits. But it raises a few questions, too. For example, a mature virtual infrastructure will cross many traditionally separate internal groups of employees, such as those responsible for servers, networks, storage and security. So a virtualization project is likely to have an impact on organizational structure and responsibilities. Therefore, getting executive commitment and support, particularly from chief financial executives able to see the big-picture savings potential of going virtual, is critical to organization-wide buy-in and cooperation.

There are basic technical and logistical questions to consider when considering a virtualization plan:

- How do you calculate the appropriate ratio to use when consolidating physical servers and other devices into software-based virtual machines (VMs)?
- What is the impact of VMs on current disaster recovery and high-availability plans?
- Do security requirements shift in a virtual environment?
- How might software licensing/pricing models, power requirements and patch management processes need to change?

Preparing a well-documented business case based on an assessment of the current environment (see AT&T paper, "Virtualization: An Overview") will help answer some of these questions. In your assessment and business case, you'll need to calculate the physical host metrics needed – such as processor type and speed; RAM amount and utilization; network interface speeds and quantities; disk resources and other metrics.

The business case should also delineate the problems you expect virtualization to solve. For example, are you currently experiencing low utilization rates on servers while growing numbers of servers are becoming unwieldy to house and manage? Going through the exercise of calculating the hard benefits will help you answer some of these questions and drive acceptance and adoption of the virtualization project throughout the organization.

Let's also take a look at each of these virtualization questions in a bit more detail.

Determining Consolidation Ratios

A primary benefit of virtualization is optimizing the use of physical equipment, such as servers, rather than letting them run underutilized much of the time and wasting money. However, it's best to avoid extremely high consolidation ratios that push physical hosts to near 100% utilization. Rather, data center managers should leave some wiggle room for moving VMs around in the event that there is a planned or unplanned outage. About 65% to 75% utilization is generally a good average to shoot for to make virtualization worthwhile but safe.

Something to consider when calculating maximum consolidation ratios is moving off of small-form-factor blade servers and onto larger chassis, the enclosures for the server blades, to allow for greater consolidation density. An approximate maximum consolidation ratio of a 1U blade server is 6:1, for example, where a maximum consolidation ratio of a 4U chassis is 30:1. So consolidation ratios might be bigger, and savings greater, with larger chassis.

Note, too, that the deployment of emerging applications, such as real-time presence and video streaming services, might require greater network capacity for shuttling communications quickly among VMs. These applications might also lower the consolidation ratio that can be achieved compared to an existing data center environment without these applications in operation.

High-Availability Implications of VMs

The wiggle room afforded by retaining a spare 25% to 35% capacity helps address failover and disaster recovery when determining the appropriate number of VMs to consolidate onto a single physical host. Fortunately, the failure of a given VM on a physical host will not affect the operation of another VM on that host, because each VM has its own isolated set of compute, memory and power resources. So a VM could fail over to another VM on the same physical host or, in the case below, to a VM on a separate physical server.

Physical Redundancy

Virtualization, again, reduces the number of physical hosts in operation by consolidating multiple VMs onto a single physical machine. The greater the consolidation ratio, then, the greater the impact will be if a single physical machine should go offline.

For example, if a single server in the past hosted a single application and served a handful of users, only that one application and those users would be affected if the server went offline. If you put 10 applications and support for hundreds of users on a single physical device, however, and the host becomes unavailable, the failure affects many more applications and users, and has greater impact throughout the organization.

You need to have a disaster recovery plan in place to address that issue. For example, a given physical host might be configured to fail over to one or more other physical hosts. Traditional high-availability configurations often require a 1:1 ratio of primary device to backup device. A virtualized environment, however, could be configured such that a physical server can fail over to one of a set of backup servers, determined, in part, by the load balancing equipment that front ends the physical servers. This allows, then, for a many-to-one backup-to-primary configuration ratio, which increases service availability (see Figure 1).

Security in a Virtual Environment

Moving to a virtualized data center means moving to two-tier security. The physical hosts must be secured, as always, and, now, so do the VMs. In unvirtualized data centers, most security is provided by special-purpose physical appliances that protect static resources, such as servers with unchanging IP addresses. In a virtual environment, however, enterprise security teams will need to deploy security tools that account for devices that might change addresses frequently in response to changing workloads.

Partitioning

Partitions should be set up to segregate a particular type of virtualized server farm from another, such as database servers from Web servers. Otherwise, malware aimed at one server farm might make its way to the other servers.

One-to-Many Failover in a Virtual Environment

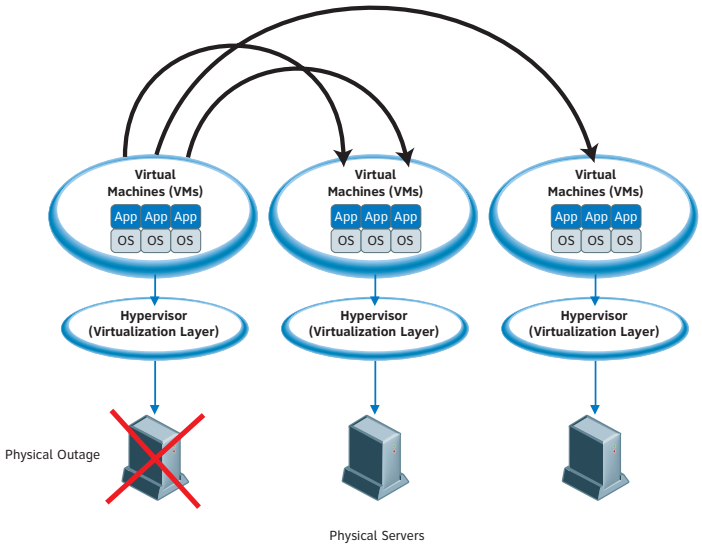

Figure 1 illustrates each VM failing over to the next best VM on other physical servers based on current load and available resources.

One approach to partitioning is to use the concept of a virtual security appliance embedded in a virtual switch that sits among the VMs to create a DMZ among the virtual server farms. The idea is that virtual security would be provisioned and re-provisioned in step with the server, storage and network resources, so that the security would adapt in parallel with the rest of the virtual environment.

Rogue VMs

One phenomenon to watch out for is so-called VM server sprawl, which can become both a security and management headache. In such instances, users with access privileges find it easy to dynamically set up VMs to suit their immediate requirements and end up creating large numbers of VMs that quickly proliferate. If this activity becomes common, VMs can fall out of the purview of the IT monitoring and management systems.

This is potentially problematic. Depending on the licensing and product support model the enterprise has in place, for example, rogue VMs that the IT department doesn't know about won't be licensed and thus won't be patched and updated, making them a potential security risk. One way to address this is by adjusting access privileges to meet the best-practice requirements of the virtualized data center. Build an administrative model designed for the virtualized environment that accounts for roles and responsibilities in a data center where, unlike in the past, it is likely that multiple people will be attempting to access multiple physical devices. Identifying these roles, responsibilities and access privileges is something that organizations should define and formalize as part of their assessment, planning process and business case development.

Also, it is possible to configure some virtualization management software such that it will trigger an alert each time a VM is created. You might want to integrate alerts with your enterprise management system console or use an automatic VM-discovery capability that is built in to VM-specific management tools and services.

Hypervisor Protection

The hypervisor, or virtual services layer, is the core software or firmware that maps VMs to physical resources. It sits between the device hardware and all its device drivers and the machine operating system. As such, it is fundamental to the virtualized data center. An attack at this level could have far-reaching consequences. Note that to date, there have been few, if any, actual hypervisor malware attacks reported. But, as with any operating system or application that becomes popular and widely deployed, the hacker attacks are usually not far behind.

Ask your virtualization vendor or service provider about the level of security here, as compromise of the hypervisor exposes all VMs on a single physical server to attack. One emerging method of protecting the hypervisor is through chip-level authentication at boot-up. During this process, the hypervisor reports the state of its integrity, using the industry-standard Trusted Platform Module (TPM) root-of-trust technology, to the hardware platform on which it is running. TPM is typically a microcontroller that securely stores passwords and other information that uniquely identify systems, software and processes. Prominent chipmakers support the TPM root of trust, which, among other things, checks the hypervisor's hash value, or cryptographic record of the system's approved configuration. If the hash value has changed and the system attempts a reboot, the root of trust will not allow it.

Impact on Physical Infrastructure

As implied in the introduction, virtualization will carry performance implications for CPU, memory, networks and storage. In order to calculate them, the maturity of enterprise performance management and monitoring systems must develop in step with the virtual infrastructure.

Resources and Capacity

If a given enterprise traditionally has been operating its physical servers and other devices at very low utilization, monitoring their performance and resource consumption may not have been a priority or even necessary up until now. With virtualization's higher utilization rates, however, tools are required for capacity planning assessments and ongoing capacity monitoring and management that account for seasonal spikes and other peak processing periods. So as the enterprise assesses its server environment to determine what capacity and resources the virtual environment will require, it should also review the capacity planning and monitoring tools it has and whether they will be up to the task in the virtual world.

Power

Migrating physical servers into a virtual environment will reduce overall power and cooling requirements. Note, though, that power requirements will be denser; more power will be required to support multiple VMs on a given physical server. In the interim, while enterprises build their virtual infrastructures, they will need to keep the physical infrastructure in place. During the migration period, enterprises will require enough power, cooling and rack space for both environments.

Time Services

Clients and servers must agree on time of day so that files can be properly synchronized. Timing services are also important to system security. For example, computers connected to the Internet must keep accurate time for evidence gathering in the case of a system break-in. Encryption and authentication protocols also require accurate time on both server and clients. For auditing and accounting, many corporate governance mandates require a log of who changed what file at what time.

There are two options for setting the timing of VMs. The first is that the physical machine supporting multiple VMs synchronizes time from a traditional source (such as an internal time server, or the U.S. Navy time clock), and the VMs, in turn, sync to the time of the physical host. Alternatively, VM timing could be synchronized with the Active

Directory component of the Microsoft Windows OS, whereby AD domain controllers set VM time.

Management and Compliance

As part of the upfront assessment mentioned in the introduction, operational procedures will need to be reviewed in the context of a virtual environment and possibly adjusted.

Software Licensing

An audit of the organization's software licensing policies is in order in planning a virtualization project. Different OS and application software suppliers charge for software licenses based on different parameters, and those parameters could become skewed in the virtual environment. For example, some application suppliers issue a license per core. A quad-core server, which has four cores in a single microprocessor on a common chip, would require four licenses. Others charge based the number of microprocessors, which in this case would require one license.

Still others charge by the number of instances of an OS. Depending on whether you use the bare metal method or hosting method (see AT&T paper, "Virtualization: An Overview"), you may be charged for many more OS instances per server. In the bare metal method, the OS is generally embedded in the hypervisor software, so you'll need to find out from your virtualization software/firmware vendor or service provider how the OS licensing works. For example, does one hypervisor license with the virtualization provider cover all the embedded OS instances?

Auditing Procedures

The enterprise will likely need new operational procedures for tracking user access and networked file management to ensure that the virtualized environment is not out of compliance with any relevant compliance mandates (for example, Sarbanes-Oxley or the Health Insurance Portability and Accountability Act). Compliance mandates might also serve as a partial guide for how to sequester, or partition, VMs and their associated security. For example, highly sensitive applications and data could be segmented from each other and from non-sensitive applications and data.

Patch Management

Patching is equally critical on physical and virtual devices. The challenge is the number of OS instances. There is generally one hypervisor per physical host, so any changes to the hypervisor software or firmware would require a single patch. However, if there are 30 instances of one or more OSs on that one host, that situation could require the same amount of patching as if patching them on disparate servers, depending on how the hypervisor is set up.

It is possible that patching will get easier with virtualization, given that the hypervisor is a homogeneous layer between the OS and software. If the hypervisor runs the OS(s) in an embedded fashion, this might alleviate the dependence on the OS supplier and allow just the hypervisor to be patched once to cover changes both to the hypervisor and to the OS(s). In such instances, the enterprises would really be relying on a new OS partner – the hypervisor supplier – for delivering OS patches and updates.

Conclusion

Virtualization involves turning a number of physical hardware computing and networking devices into software and loading them onto a common high-powered hardware platform. When deployed in enterprise data centers, it affords a number of cost and energy-conserving benefits. Still, organizations can hardly go into a virtualization project without assessing how such a project will impact traditional operations, both technically and organizationally.

Getting executive support, particularly from corporate financial executives, is critical for success. Administrative rights and models will have to be reviewed, along with roles and responsibilities, as individuals traditionally in charge of one or two applications might require access to a number of physical machines in the new scenario. In addition, enterprises must make sure that their capacity planning and monitoring tools are up to the job of regularly assessing utilization and application performance in a virtualized environment.

From there, enterprises can calculate the software-to-hardware ratios that allow enough spare cycles to accommodate a potential VM failover while still yielding the savings ROI that makes the project worth the organization's while. Data center security strategies must account for both the physical devices as well as the virtual devices. They must also consider that the hypervisor software or firmware selected plays a pivotal role in virtualization and work with the virtualization vendor or service provider (depending on whether they bring projects in-house or outsource them to a professional services organization) to determine the level of risk and protection needed at the hypervisor layer.

Virtualization will also have some impact on the physical infrastructure in terms of the compute power needed in the physical devices supporting multiple VMs, as well as in the network capacity required to shuttle requests among virtual and physical machines.

These are a lot of considerations, and it can behoove enterprises to engage with a vendor or services partner upfront to make sure that all the variables are taken into account before the virtualization project is in progress.

References

1. "U" denotes the height of network or computing equipment mounted in 19-inch-wide or 23-inch-wide rack units, or RUs, in an equipment room. One RU is 44.45 mm (1.75 in.) high. Equipment that fits into one RU is commonly designated as "1U"; similarly, equipment taking up 2 RUs are "2U" and so on.

For more information contact an AT&T Representative or visit www.att.com/business.

12/08/08 AB-1394