Virtualization: An Overview

Executive Summary

Improving resource utilization through virtualizing IT infrastructures is becoming a priority for many enterprises. A successful deployment requires up front preparation to determine the appropriate infrastructure components and architecture. This paper explains the basic concepts behind virtualization, potential benefits and key decisions associated with a virtualization project and how to get started with a virtualization assessment.

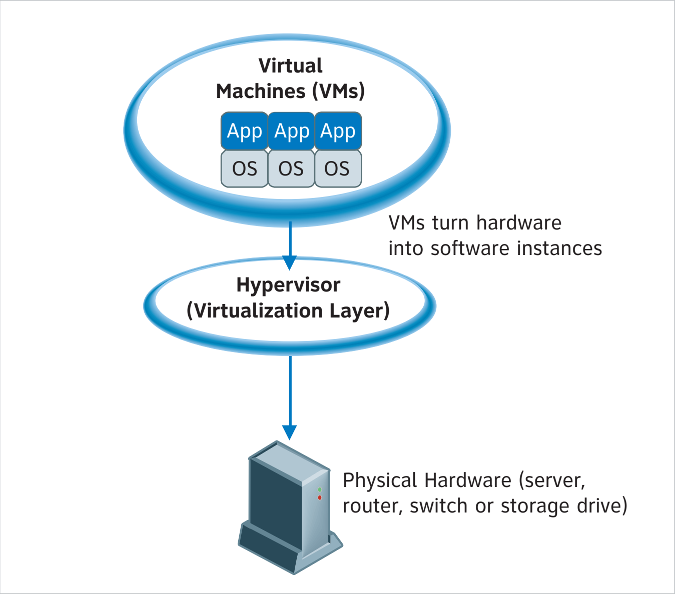
Introduction

Virtualization projects are the focus of many IT professionals who are trying to consolidate servers or data centers, decrease costs and launch successful “green” conservation initiatives. Virtualizing IT resources can be thought of as squeezing an enterprise’s computer processing power, memory, network bandwidth and storage capacity onto the smallest number of hardware platforms possible and then apportioning those resources to operating systems and applications on a time-sharing basis.

This approach aims to make the most efficient possible use of IT resources. It differs from historical computing and networking models, which have typically involved inextricably binding a given software application or service to a specific operating system (OS), which, in turn, has been developed to run on a particular hardware platform. By contrast, virtualization decouples these components, making them available from a common resource pool. In this respect, virtualization prevents IT departments from having to worry about the particular hardware or software platforms installed as they deploy additional services. The decoupling and optimization of these components is possible whether you are virtualizing servers, desktops, applications, storage devices or networks.

To virtualize some or all of a computing infrastructure’s resources, IT departments require special virtualization software, firmware or a third-party service that makes use of virtualization software or firmware. This software/firmware component, called the hypervisor or the virtualization layer, performs the mapping between virtual and physical resources. It is what enables the various resources to be decoupled, then aggregated and dispensed, irrespective of the underlying hardware and, in some cases, the software OS. In effect, the hypervisor takes over hardware management from the OS. In addition to the hypervisor virtualization technology, the organization overseeing the virtualization project requires a virtualization management tool – which might be procured from the same or a different supplier – to set up and manage virtual devices and policies.

Virtualization Model



Why Virtualize?

One key reason why IT organizations are considering virtualization of some or all of their computing infrastructures is that the technology helps them to derive the biggest bang out of their computing buck. Consider, for example, the case of the server infrastructure. By bumping up the utilization of one or more application servers from 15% each to 85% each, enterprises can eliminate a significant number of physical servers. They simply consolidate what would have been multiple physical servers onto one machine running a number of virtual, or logically separate, servers.

Optimizing the Virtualization Project

Many organizations that have gotten started with virtualization projects have taken an overly conservative approach, sometimes leaving significant savings on the table. A large East Coast publishing company, for example, initially virtualized its server infrastructure for an 8:1 consolidation with the help of its primary hardware supplier.

AT&T Consulting, which offers virtualization services, then entered the scene and found operational opportunities that allowed the publisher to boost its server consolidation ratio to 28:1. This resulted in approximately a $5.6 million savings in hardware cost avoidance and 21,000 kilowatt hours per year in power savings, compared to the company’s initial physical infrastructure.

AT&T differentiates itself by its operational assessments and by assisting customers with building out the “people and processes” aspects of virtualization. Its Virtual Infrastructure Operational Review specifically assists organizations that have already rolled out a virtual infrastructure in conducting a detailed operational assessment to help ensure that the business and operations processes have been properly aligned to support the virtual infrastructure.

Historically, there has been a 1:1 ratio of server to application, because specific interfaces have bound them together. This has left many CPU cycles sitting unused much of the time, dedicated to a particular application even when there are no requests in progress for that application. Now, IT departments can run more than one OS and application or service on a single physical machine and access each through separate windows on their virtualization management console.

Hardware resources can be applied, dynamically, to whichever application(s) needs them if the appropriate resource allocation tools are in place. This setup delivers several cost and productivity benefits that are described below.

Lower Expenses

Reducing the number of physical devices needed to get the job done naturally lowers capital costs. It also decreases operating expenses by leaving IT departments with fewer physical devices to manage. In addition, packing more cycles into fewer pieces of hardware consumes less energy overall and requires less floor space, a boon to today's widespread enterprise efforts to consolidate data centers into fewer and fewer locations and to embark on "green IT" initiatives.

Business Continuity

Virtualization contributes to higher levels of business continuity in a couple of ways. For example, with the decoupling of software applications, operating systems and hardware platforms, fewer redundant physical devices are needed to serve primary machines. Traditional high-availability configurations often require a 1:1 ratio of primary device to backup device in addition to the 1:1 software-to-hardware ratio discussed earlier. In the virtualized environment, however, multiple servers can fail over to a set of backup servers. This allows, then, for a many-to-one backup-to-primary configuration ratio, which increases service availability.

High Availability

Virtual devices are completely isolated from one another, as though they were running on different hardware, which alleviates downtime during patching and updates. This means that hitless changes can be made to one virtual device without affecting others sharing the same hardware; in other words, changes can be made in a production environment without having to schedule downtime.

Fast Installation

Similarly, virtual devices allow for much faster installation of new server applications or router/switch software services, because the IT department no longer has to purchase additional equipment that can take days or weeks to order, arrive and set up. Rather, IT administrators simply configure a new virtual server, desktop, router, switch or storage drive using the special virtualization management software tool mentioned. This process generally consists of clicking on an existing image, copying it and pasting it, thereby slashing the setup times to minutes. In addition, in the case of server and network virtualization, discussed later, enterprises can set policies for how servers and networks address competing resource requests, such that resources get allocated to high-priority applications first.

Corporate Governance

A number of corporate governance mandates, including Sarbanes-Oxley, Gramm Leach Bliley, Healthcare Information Portability and Accounting Act (HIPAA) and others, have placed stronger privacy, security and auditing requirements on organizations. This has led enterprises to a more consolidated computing and networking infrastructure. It is inherently more manageable to set and enforce policies and configure software from a central, common console than in a distributed fashion, which makes it difficult to keep software versions synchronized. With fewer data centers and less real estate available for housing the same or more computing and networking cycles, going virtual helps support the higher-level management, security and tracking that the tighter mandates require, in a cost-effective way.

What Can You Virtualize?

As noted, there are several computing infrastructure components that can be virtualized. When aggregating computers and their attached networks and storage into a unified pool of IT resources, virtualization makes one thing look like something else. For example:

- Servers. From an access and management perspective, a single physical server would appear to be multiple servers, often called virtual servers or virtual machines (VMs).
- Desktops. Similar to server virtualization, desktop virtualization can mean one of two things. First, it is possible for users to run multiple desktop OSs – such as the Apple Mac OS and Microsoft Windows XP or Vista OS – on the same computing device. Second, and perhaps more important to the corporate IT department, desktop virtualization also can allow a user's own data and services to reside on a computer shared by others' data and services. The use of a software component called a connection broker enables the user to connect to his or her virtual desktop – to the data and services associated with that user – in many different ways, such as through a thin client, existing desktop, laptop and/or Remote Desktop Protocol (RDP).

Desktop virtualization has both security and administrative cost advantages to the IT department. First, the user's client device becomes just an access device, rather than housing potentially sensitive data that could be easily compromised by loss or theft. Virtual desktops will likely share server hardware with virtual servers, potentially resulting in 30 to 40 desktops being consolidated onto a single piece of computing hardware to purchase and manage. Also, today's dual-core and multi-core integrated circuit designs allow multiple processors to coexist on the same computing chip, which renders server hardware still more powerful. As a result, IT departments gain the ability to fit increasingly more VMs, which can be virtual desktops, virtual servers or a mix of the two, onto small (1U – 4U[1]) but powerful computer platforms for better economics.

- Storage. One physical storage drive appears as multiple isolated sub drives or virtual drives. In other words, using separate windows on a common management console, IT administrators can treat each virtual drive as though it were a distinct physical drive.
- Applications. When virtualized, applications written for one OS environment can execute in another operating environment for improved application compatibility and manageability. This happens because applications are encapsulated from the underlying operating system on which they are executed. Operations are redirected to the appropriate operating system.
- Networks. In a network, a single physical router might support multiple, partitioned IP addresses to create virtual routers. Similarly, one physical Ethernet switch might support multiple media access control (MAC) addresses to create virtual switches. Again, a single piece of physical hardware can be divided up into multiple virtual routers or switches to reduce expenses.

In addition, network bandwidth can be virtually partitioned for privacy, security and economies of scale through the use of virtual LANs (VLANs) and virtual private networks (VPNs). See box for a detailed discussion on VLANs and VPNs.

The goals of virtual networks – whether local, wide-area, switched, or routed – are the following:

- To optimize the use of aggregate capacity, or network bandwidth, by sharing it among user groups and applications, rather than building and managing separate networks for separate applications
- To keep traffic private, despite the shared nature of the network

VLANs and VPNs

Virtualization in networks enables network operators to carve up an aggregate pool of bandwidth for sharing among different user groups, business units, applications or enterprises. Many IT departments are familiar with the concept of a virtual LAN, or VLAN, which logically separates internal switched Ethernet traffic that traverses the same physical cabling into partitioned groups.

Users or applications are allowed onto each VLAN by a corporate policy, which places them in a common logical group. The network administrator assigns every member of the logical group a VLAN number, known in standards parlance as an 802.1Q tag, as well as associated access credentials for accessing that VLAN. Users in one VLAN cannot "see" traffic in another VLAN, even though they generate and receive traffic that shares the same physical cabling with the other VLANs.

The same concept is applied in wide-area switched or routed networks using virtual private network, or VPN, technology. This partitioning can be accomplished in several ways.

Virtual Circuits

One method is by creating virtual circuits in Layer 2 frame relay and ATM networks. Each virtual circuit is assigned a virtual channel identifier (VCI) number and each type of traffic allowed onto that virtual circuit is assigned that VCI number. This is conceptually similar to the VLAN, albeit across a WAN infrastructure.

In metro or WAN Ethernet services, which are also inherently Layer 2 services, a service provider will often add a second tag to a customer's internal Ethernet 802.1Q VLAN tags. The additional tag creates the VPN across the WAN, keeping that customer's traffic segregated from other WAN traffic. The service provider preserves the original VLAN tags as well, thereby retaining the customer's internally designated user groups and associated access rights across the WAN. This re-tagging can be thought of as creating a local VPN within a VPN, or, more accurately, a VLAN within a VPN. This virtual network service is sometimes referred to by the standards-based technology it uses, which is called 802.1Q-in-Q, or just Q-in-Q.

Encrypted Tunnels

In Layer 3 routed networks, encryption is often used to separate one user's organization from another and also to separate a single organization's own internal VLANs from one another across the WAN. Encryption creates protected "tunnels" through a shared IP network to keep traffic separate by appending an encrypted IP address packet to the "real" address packet and stripping it off at the other end before delivery. The encrypted streams of traffic cannot interpret one another as they traverse the WAN.

- To further optimize bandwidth use by time-sharing and enabling dynamic resource allocation to user groups and/or applications as they need it
- To allocate resources to high-priority traffic based on the enterprise's policies

Together with application virtualization (described above), network virtualization contributes to the emerging trend toward software as a service, or SaaS. In the SaaS model, businesses elect to use application software capabilities hosted on a service provider's remote server, accessed by way of the network, for a monthly fee. This alleviates the organization from having to purchase or develop the software themselves, from buying and installing the hardware to support it on premises and from having to manage and maintain it.

VMs, whether they are servers or desktops – or even storage drives, routers or switches – each have their own CPU, memory and power source. As such, they are completely isolated from each other logically, though they share underlying physical hardware. The various types of virtualization described together create a virtual infrastructure. It is possible to virtualize your entire infrastructure or selected components of it.

Getting Started: Conduct an Assessment

Getting started with virtualization requires some upfront decision-making. First, determine which components of your computing infrastructure you would like to virtualize. Do you want to virtualize just your server infrastructure? Desktops, too? Storage resources? Network devices and network bandwidth?

Answering these questions requires that you gather current statistics about the utilization of each of these resources. To do so, you can use an in-house tool that you have already developed or a third-party tool from a virtualization specialization company or service.

Using a third-party service often provides a more objective view of the environment than an in-house approach. Also, third parties often are able to tap data that are difficult for people working within the IT organization to see because of internal restrictions on their access rights.

It is possible that bringing the project in-house and using existing IT staff could be a lower-cost alternative. Tackling the job internally could prolong the virtualization process because IT personnel already have other job responsibilities. Prolonging the virtualization project would also delay its associated return on investment.

Whatever your measurement method, it is those resources that are significantly underutilized that are the best candidates for virtualization and cost savings. For example, it is not uncommon to see a mere 3% to 6% utilization on some very robust servers. Creating VMs that together get 60% to 70% usage out of one of those physical servers will reduce the expense and management costs associated with extra, unnecessary servers, while still leaving enough headroom on the physical device to accommodate load spikes and failover traffic.

You will want to gather your statistics for a length of time that allows you to capture both the highest and lowest utilization markers, so the monitoring time period will differ from organization to organization. Businesses that generally see the same peaks and valleys month after month, for example, might gather usage stats for a 30-day period. More seasonal businesses, such as retailers that see peaks during holiday seasons, might deploy the tool for 60 to 90 days.

Software or Service?

You may elect to "buy" some or all of your virtualization in the form of a service or to "build" your virtualization architecture yourself. Virtualization services are becoming commonly available and can fall into a couple of different models. For example, server virtualization might be offered as a component, or subset, of a more broadly encompassing set of managed IT services. In such a model, enterprises outsource their server infrastructures, often for a flat fee. Their infrastructures are maintained on the service providers' premises and operated according to the enterprise's own policies for dynamic resource allocation priorities. In this scenario, it is the managed IT services company that deploys the virtualization hypervisor and management tool in its own infrastructure. The provider dedicates some of that virtualized infrastructure to the enterprise customer, which would use its portion of the infrastructure as if it were its own data center.

One form that outsourced virtualization services can take is utility computing. Virtualization and utility computing are similar in concept, in that resources are time-shared and serve requests on an as-needed basis. Utility computing, however implies a usage-centric pricing model. This model can be used in the context of a service provided by a third party, or, similarly, by an in-house IT department. In the latter scenario, the IT department could set up a utility environment whereby charge-backs to various business units and departments could also be on a usage basis. The usage model might be paying for each increment of resource as it is used or a fixed fee per VM.

If you decide to build your own infrastructure, you must decide on a type of virtualization architecture, or the basic technological approach you prefer. There are two main types used for servers, for example, and they have to do with how the OS is affected by the hypervisor:

- Bare metal method. In this architecture, used for data center rollouts, the hypervisor is a microkernel, or small OS, in its own right. This method involves loading the hypervisor, then the VM, onto the hardware. The VM contains the OS and applications, and the hypervisor manages the resource requirements of the VMs. If using a product or service based on the latest virtual infrastructure software, a single patch to the hypervisor could replace the previously separate and frequent patches required to OSs and applications. Patching multiple components concurrently is highly recommended for the enterprise environment to keep OpEx costs down and software versions synchronized.

- Hosted virtualization. In this architecture, there is an interface between the hypervisor and the application, and time-sharing takes place on the processors below. The original software developer's OS still controls the underlying hardware and hosts the VM applications. This architecture is recommended for small implementations, such as testing virtualization out on a spare server or other resource, because the original OSs have not been customized for virtualization.

Once you have decided on an architecture, you need to choose either a software or firmware VM platform. This choice is usually vendor-dependent; some VM vendors provide their products in software and others use firmware, but most don't offer a choice. From there, you can see which suppliers or virtualization products fall into which camp and begin evaluating products and services accordingly.

Conclusion

There are compelling cost- and energy-saving reasons to virtualize some or all of a computing infrastructure. By employing a homogeneous VM software or service layer between traditional software and hardware platforms, hardware resources can be more efficiently utilized on a time-sharing basis to consolidate devices, conserve energy and real estate, boost business continuity and serve high-priority applications and users. There are many aspects to virtualization, which can be taken advantage of in the form of a service or by deploying an internal virtualization infrastructure in the enterprise's own data center(s).

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.

at&t
Your world. Delivered.