Realtime publishers

The Definitive Guide™ To

Virtual Platform Management

sponsored by

ca technologies

Anil Desai

Copyright Statement

Chapter 5: Managing Heterogeneous Virtualization Environments

One of the many benefits of working in the IT industry is the ability to choose from a wide variety of different technologies. In the areas of hardware, software, and infrastructure, there are literally hundreds of options from which you can choose. In some cases, one solution or approach might be ideal. In most cases, there will be various pros and cons related to the decision. These tradeoffs often require IT environments to move to take the approach of supporting multiple different systems, applications, and services. The benefits include the ability to choose the best solution for various different problems.

But there is a downside: Managing heterogeneous environments can be complicated and time-consuming. Many areas of IT architectures lack standardization, and vendors have very different ways of administering their products. This places additional burdens on systems administrators and other IT staff that must manage this complexity while still ensuring that business needs are being met. The primary goals are availability, reliability, and performance. Specific concerns include capacity planning, performance monitoring, deployment, configuration management, and conforming to organizational policies and processes.

Virtualization technology provides both solutions and challenges related to managing disparate systems. On the positive side, the use of virtualization can help dramatically improve hardware resource utilization and create an abstraction layer between a workload and the physical hardware on which it is running. The end result is improved portability of workloads and access to a wide array of platforms that can improve operations. IT organizations have a plethora of different types of virtualization platforms from which to choose. Unfortunately, due to the rapid rise in popularity of this technology, there is a lack of standardization in management tools and techniques. Systems administrators often have to learn several platforms and technologies to get their jobs done. Clearly, there is room for improvement.

The focus of this chapter is on identifying the various types of systems—both physical and virtual—that are often found in enterprise IT environments. We'll start with details about the various platform options that are currently available. Details will include the general benefits and drawbacks of each type of system. We'll then use this information in identifying common environment management issues. And finally we'll look at several best practices for managing heterogeneous environments.

Heterogeneous Virtualization Platforms

There are several virtualization platforms available, and organizations might have good reasons to use different ones for different workloads. For example, some virtualization platforms (such as Microsoft Virtual PC and VMware Workstation) have been designed to run primarily on desktop OSs. On the server side, products such as VMware ESX Server, XEN, and Microsoft Virtual Server provide good options. In addition, as new platforms and versions are introduced, IT departments must be able to determine which approach works best for a given workload.

Chapter 3 looked at the different types of virtualization solutions that are available in the IT marketplace. The numbers and types of products from which IT organizations can choose are extensive. Each type of solution provides a combination of benefits and drawbacks. Figure 5.1 shows a summary of major types of virtualization implementations.

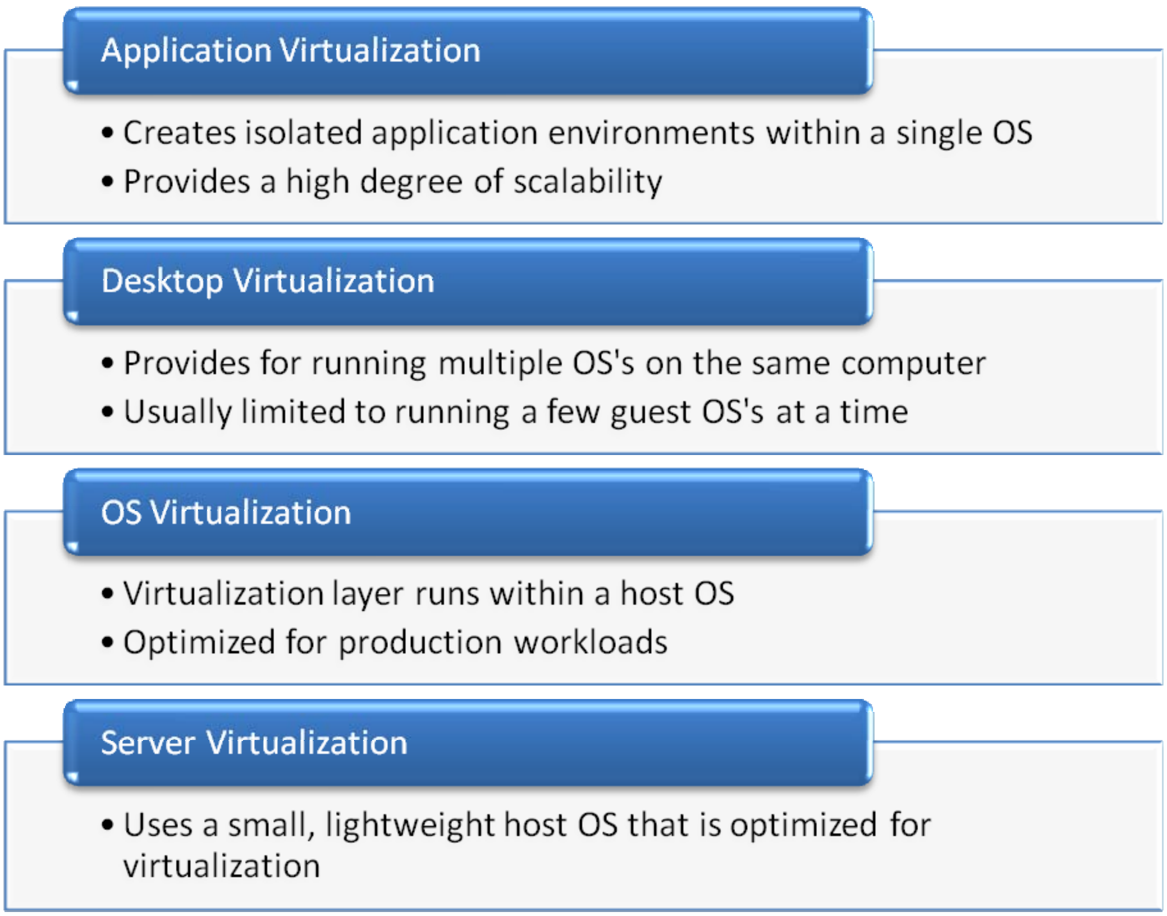

Figure 5.1: An overview of different virtualization approaches.

Each of these implementation methods comes with its own pros and cons. Therefore, IT organizations will often choose to support more than one platform, leading to a heterogeneous virtual environment. Next, let's take a brief look at each approach, along with a summary of the types of products that are currently available.

> It is important to note that this list is not intended to be exhaustive, and it is very likely that new applications and technologies will become available in the near future.

Application Virtualization

When systems administrators think of the term virtualization, they most often think of the situation in which multiple independent OSs are running concurrently on the same physical hardware. In some cases, however, it might be unnecessary to run several OS instances. For example, if an administrator wants to run multiple instances of a single-user office productivity suite on the same computer, virtualizing the file system, registry, and other settings might be a more efficient method.

Often, application virtualization can be far more scalable than other approaches. Due to the fact that only a single host OS is running at a particular time, it's not uncommon for systems to support dozens or even hundreds of independent operating environments on the same computer. Examples of application virtualization products include Citrix Application Virtualization, Microsoft SoftGrid (formerly known as Softricity), and SWSoft's Virtuozzo.

The primary drawback of application virtualization is that of compatibility. In some cases, applications may need to be modified to run in their independent environments. Also, as all of the environments are running on the same host OS, some types of differences aren't possible. For example, every instance of the application must run on the same OS with the same patches installed.

Desktop Virtualization Platforms

Desktop virtualization platforms are designed to be quick and easy to deploy and manage. They're designed for end users and are usually available either for free or at a nominal price. A primary benefit of these products is the ability to quickly install the product and start creating new virtual machines. Usually, this can be accomplished with limited knowledge of standard IT deployment practices. Product examples include Parallels Workstation, VMware Workstation, and Microsoft Virtual PC (see Figure 5.2). These products install as client applications and provide a simplified user interface for starting, stopping, and interacting with virtual machines.

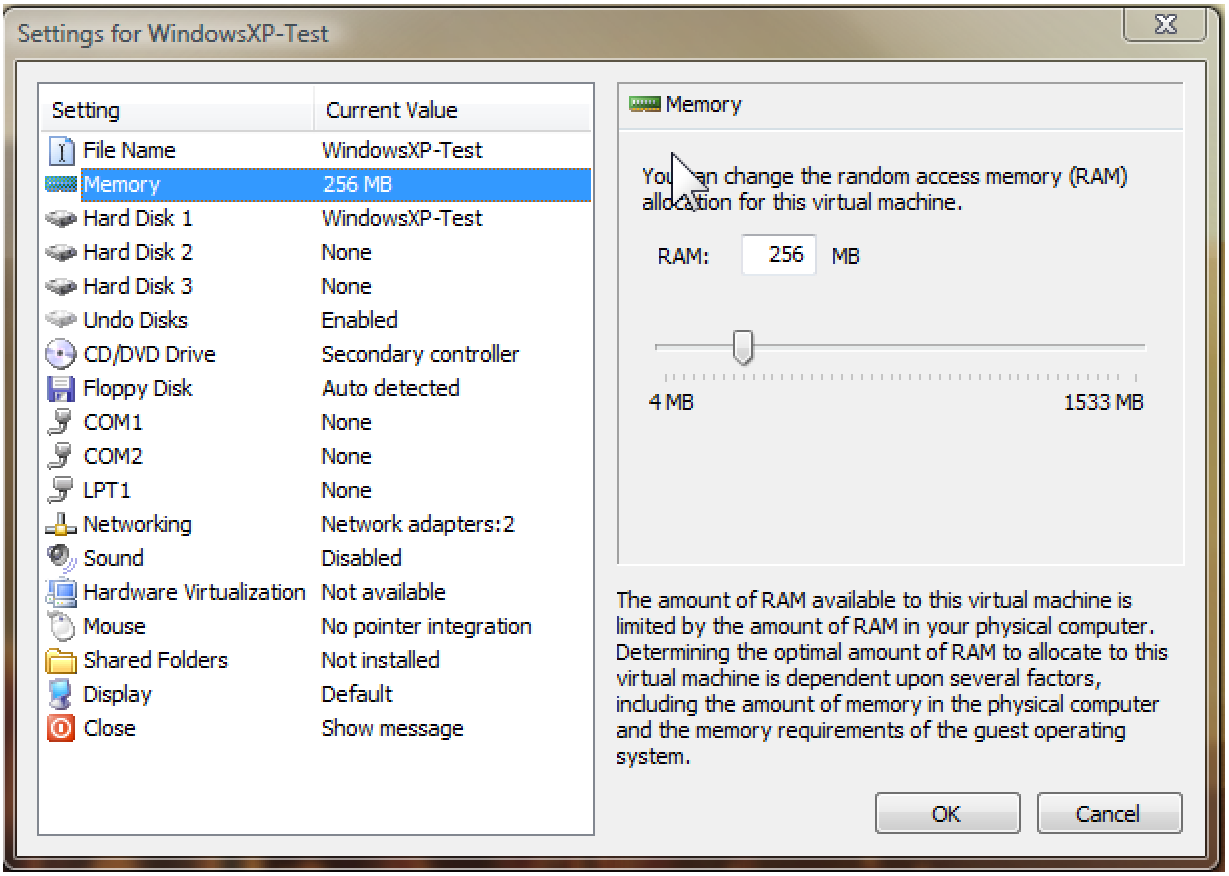

Figure 5.2: Configuring a virtual machine in Microsoft Virtual PC.

The primary tradeoffs with desktop virtualization are related to a lack of scalability and manageability issues. Most of these solutions are not designed with centralized management in mind and often require additional effort to be deployed in production environments. For example, configuration of the virtualization layer is managed independently on each system, and administrators will need to connect to each computer to make changes. Additionally, users will often have the ability to perform advanced tasks—such as placing a new virtual machine on a production network—without any IT oversight. Still, this solution is a convenient option for users such as software developers and QA staff that want to get up and running with one or a few virtual machines quickly.

OS Virtualization

For running multiple independent production workloads on the same hardware, IT departments commonly use OS virtualization. The term OS virtualization is used here to refer to the use of a virtualization layer that runs atop a physical computer's OS. The virtualization layer creates multiple independent virtual machine environments, each of which can host a separate guest OS.

Like desktop virtualization, this method allows for running multiple isolated virtual computing environments. However, OS virtualization solutions have been designed to support mission-critical applications. Therefore, they include enhanced scalability and support for larger hardware platforms. Most solutions offer support for multiple host and guest CPUs as well as support for large memory. Their architecture is designed to support many concurrent virtual machines. Other enterprise features include security features, centralized management, and support for high-availability features.

Examples of OS virtualization solutions include the XEN open source platform, VMware Workstation and Server, and Microsoft Virtual Server. Although the features and capabilities of the products differ, the primary goal is to provide IT departments with the features they need to manage production workloads.

There are a few potential drawbacks of OS virtualization. The first is that scalability is often limited based on the hardware configuration of the host server. Memory, CPU, and disk limitations often place a ceiling on the number of different OSs that can run simultaneously. Additionally, the presence of a host OS can cause overhead, which uses resources and can reduce overall performance.

Server Virtualization

It is becoming increasingly common for systems administrators to deploy computers solely for the purpose of running virtual machines. In these configurations, the presence of a full-featured host OS may be unnecessary. In fact, a host OS often adds unnecessary load to the computer. In these cases, server virtualization may be a more appropriate solution. The term server virtualization is used here to describe an architecture in which a physical computer runs a small OS that is dedicated to and optimized for virtualization. It generally involves the use of a thin, dedicated virtualization layer that is able to directly (or closely) interact with the underlying hardware. Some implementations may rely on remote computers for performing management functions, while some might include basic systems management tools in a small layer that runs alongside the host virtualization platform.

Examples of server virtualization solutions include VMware ESX Server and Virtual Infrastructure and Microsoft's upcoming Windows Server Virtualization technology. The primary benefits of this approach include increased scalability and performance due to closer interactions with the underlying hardware. The primary drawbacks include a potential lack of manageability (because a complete host OS is usually not available). For example, standard GUI-based administration tools and features may not be available for performing common tasks. This increases the learning curve for administrators. Additionally, server virtualization solutions may be limited to running on a certain list of supported hardware. This constraint is necessary, as some platforms require specialized drivers to be written to communicate directly with the underlying physical hardware.

Vendor-Based Solutions

So far, we have focused on a variety of different software-based approaches of implementing virtualization technology. Most of these solutions run on a wide array of hardware types, making them compatible with many existing IT assets. In addition to these methods, several vendors have provided virtualization solutions that can be used to simultaneously run multiple concurrent workloads. Some examples include:

- IBM Logical Partitions (LPAR)
- HP Virtual Partitions (VPAR)
- Sun Logical Domains (LDOM)

Each of these solutions provides its own features and capabilities but the overall goal is to allow for a single physical server to host many different applications and services simultaneously. IT departments that support multiple vendors will need to consider the management of these systems as part of their routine administration.

Heterogeneous Physical Platforms

Without even considering virtualization, the process of managing physical computers can be a tough one. Due to differing business and technical requirements, many organizations will find the need to support various types of hardware. This section will look at different types of physical computer assets (several of which might be used for virtualization host computers), with a focus on management issues.

Client-Side Computers

Based on sheer quantity, most organizations will support far more end-user computing devices than they will support servers. These devices are often issued to individual employees for the purpose of running corporate applications. User types range from the standard “desktop” deployment to traveling employees who require mobile support. A brief list of the types of devices that must be considered includes those shown in Figure 5.3.

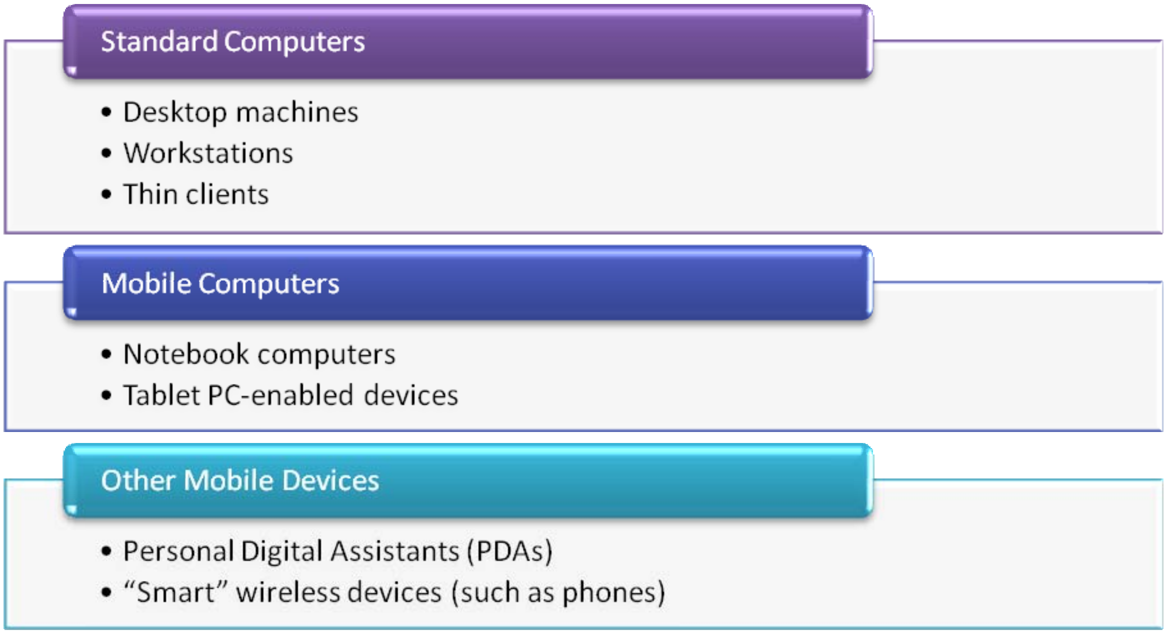

Figure 5.3: Types of client-side computing devices.

Challenges include managing the configuration of each device and ensuring security.

Rack-Mounted Servers

If there is a "standard" physical server deployment for most IT departments, it is the rack-mounted server. These computers are designed to be deployed into standard data center racks and can support fairly dense deployments. The servers themselves can range in physical size and capabilities (see Figure 5.4). Smaller configurations are often 1U servers (where a "U" is equivalent to 1.75" of rack space). They usually contain as many as two physical CPUs and support up to two direct-attached physical hard disks. These machines are often used as Web servers and applications that do not require significant resources.

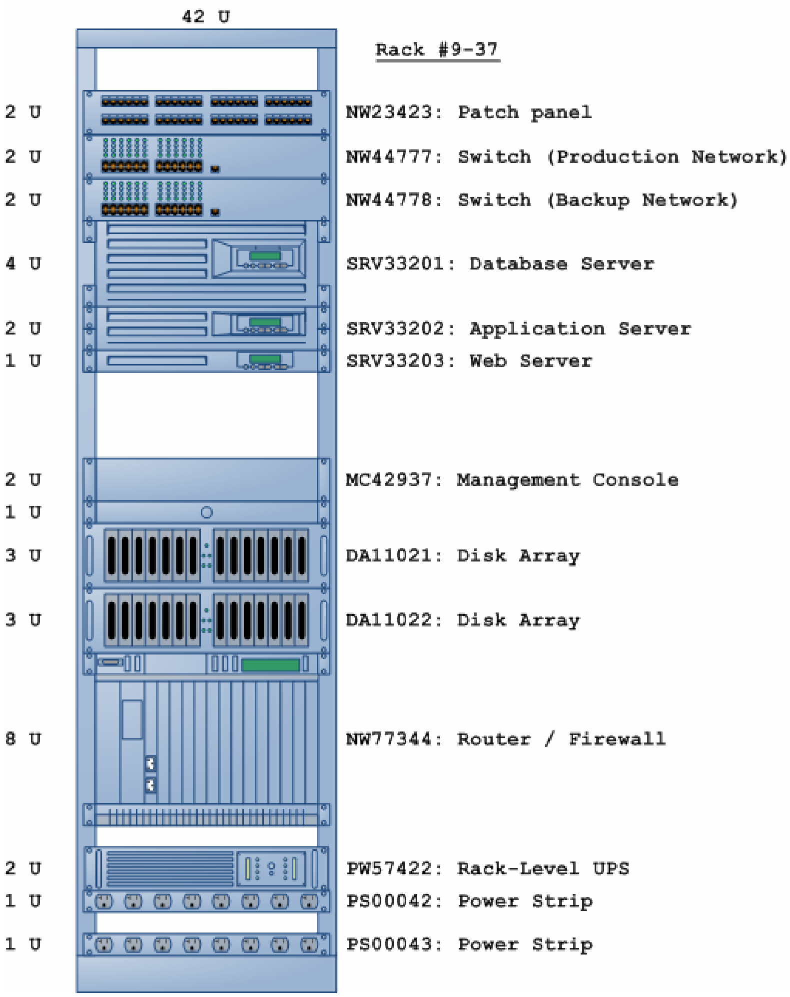

Figure 5.4: An example of a rack configuration.

2U and 4U servers take up more physical space but provide for more CPUs, more hard disks, and larger memory. Most rack-mounted servers are designed with high-availability in mind and include redundant power supplies, cooling fans, and other components. In general, rack-mounted servers are managed through their primary OSs and with enterprise management tools. Many vendors also include remote management capabilities, network-based access, and dedicated ports for performing base-level tasks.

Blade Servers

The primary constraints in many data centers are related to physical space (such as floor- and rack-space), cooling, and power. Physical space may be difficult (or impossible) to increase, so IT departments are always looking for ways to squeeze more computer resources into a limited amount of rack space. Blade servers offer a base chassis that contains slots into which blade servers can be installed. The chassis is responsible for providing power and some management functionality. The blades themselves include their own physical hardware, including CPU and memory. Some solutions also allow for direct-attached storage, while others are primarily designed to be booted from the network.

From an IT management standpoint, blade configurations often require additional administrative effort. Each blade will typically be running its own OS, so standard enterprise management tools can be used. Managing the hardware itself can be more complicated. Each blade vendor typically provides its own set of management tools and features. These must be taken into account when supporting heterogeneous systems.

Clustering Solutions

As organizations have come to place increasingly heavy loads on critical applications and services, it often becomes too difficult or too costly to meet these needs within a single physical server. In these situations, IT departments often turn to clustering solutions. A basic definition of a cluster is a group of individual physical computers that act as one logical unit. Examples of clustering can be seen in various areas, including storage, networking, and databases.

The underlying architecture of a clustering solution includes the ability to create a single logical entity that will coordinate access to physical servers. This is most often implemented at the network layer. Additionally, server clusters can provide for automatic failover and load-balancing functionality. These features allow for increased performance and high-availability for mission-critical workloads. Several vendors provide clustering solutions:

- VERITAS Cluster Servers (running on Sun Solaris, Linux, and Windows)
- IBM High Availability Cluster Multiprocessing (HACMP)
- HP Serviceguard
- Microsoft Cluster Services

From an IT management standpoint, systems administrators must have the ability to configure the cluster and to manage individual computers that are part of the cluster. Being able to visualize these relationships can be helpful in ensuring that everything is running properly.

Challenges Related to Managing Virtual Environments

Thus far, we've seen how many IT departments often have the need to support many different types of resources. This level of flexibility provides systems administrators with the ability to choose the right type of solution for a particular workload. The benefits include increased efficiency and better performance for applications and services. Unfortunately, there is a downside: Many IT departments face a difficult challenge when trying to manage a broad range of physical and virtual platforms. Before we get into the details, it will be helpful to envision a long-term goal for systems management.

There are several important characteristics of an ideal computing environment. The primary feature is the ability to treat all physical assets as a single pool of hardware resources. Regardless of the details of the underlying physical platform, CPU, memory, disk, and network resources would appear as one large seamless pool. Workloads could be easily placed in the pool, and automated management tools would automatically determine where they would best be placed. Performance would be monitored, and workloads could automatically be switched between physical and virtual environments based on current usage.

In some ways, various technologies have helped organizations move in this direction. For example, storage virtualization and associated technologies allow systems administrators to centralize hard disk resources while still providing direct access to those who need it. In other cases, the complexities of the underlying infrastructure reduce some of these goals to just wishful thinking. This section will look at some of the challenges that make managing heterogeneous environments difficult.

Virtual Machine Portability and Virtual Machine Sprawl

Deploying physical computing resources often involves numerous steps and interactions. Although the process can be time-consuming and labor-intensive, it does lend itself to certain levels of control. Almost always, cooperation between the IT department and end users is required. Details such as performance requirements and metrics must be considered, and the appropriate resources are purchased and deployed.

The situation is quite different when using virtual platforms. Unlike their physical counterparts, virtual machines can be quickly and easily moved between servers. Often, the process can be performed in a matter of minutes, and by an individual user. Although there are numerous benefits of this ability to move workloads, it creates a fluid environment that IT departments must manage. They can no longer rely on physically inspecting server racks in the data center. Often, organizations find that they are faced with a problem that is known as "virtual machine sprawl." The number and types of virtual machines that are deployed across the environment can quickly become unmanageable.

Heterogeneous Management Tools

One important consideration related to the management of heterogeneous environments is that of cross-platform administration. In the realm of network devices, the Simple Network Management Protocol (SNMP) provides a consistent way in which to communicate with a broad array of network-based devices. However, in the world of physical computers, there are no universally accepted standards that allow for managing every type of device from every vendor. For this reason, IT departments often standardize on certain types of server hardware.

In the world of virtualization, application vendors have generally created their own platform management tools. For example, administration methods for Microsoft Virtual Server and VMware ESX Server are significantly different. Although both systems primarily rely on a Web-based management system, the actual operations that must be performed are significantly different. When other platforms are added, this can quickly become a serious management issue.

Each virtualization platform comes with an associated learning curve, and systems administrators often have to log on to multiple systems to manage different groups of virtualization host servers. This lack of consistency makes it a challenge to retain control of disparate, dynamic environments.

Platform-Specific Features

Earlier, this chapter looked at the pros and cons of each type of virtualization platform. These differences can provide many advantages to IT departments that are trying to maximize the efficiency of their data center assets. One potential management challenge, however, is that virtual machines are often incompatible between different vendors' solutions. For example, Microsoft Virtual PC and Microsoft Virtual Server can, for the most part, host the same virtual machines. However, virtual machines created in VMware's products are not directly compatible with those from Microsoft (although conversion tools are available). Additionally, particular features, such as the ability to take multiple "snapshots" of a virtual machine at a specific point in time, might not be supported on all platforms. From an administration standpoint, IT departments will need to build expertise and standards related to which virtualization technologies they will support.

Resource Allocation

An important benefit of virtualization technology is the ability to dynamically move workloads based on their specific requirements. Ideally, any workload could be moved to any server or other resource at any time. In reality, it's difficult to ensure compatibility when moving applications and services between platforms. The primary reason for this is that OSs and applications are tightly coupled to the systems on which they are running. Additional details such as OS and software configuration can have a dramatic impact on overall performance of the workload. Virtualization technology helps reduce this barrier by allowing workloads to be easily moved to other physical servers.

This raises another important question: How should virtual machines best be deployed to a given set of virtual servers? And how should the virtual machines themselves be configured based on performance requirements? Often, virtual machines are given more resources than they need. For example, a rarely used virtual machine might be allocated 2.0GB of physical memory. Even though the virtual machine is not using this resource, it is unavailable for use by other workloads. The end result is less-than-ideal utilization of resources.

Visualizing the IT Landscape

The flexibility of choosing from various platforms—including virtualization, physical servers, and clustering—can make it difficult for IT staff to determine the exact configuration of a workload. Often, it's difficult to detect the relationships between virtual machines and host computers. The same is true for clusters: Does a particular network name refer to an individual server or a logical cluster of machines? Many enterprise network management tools are unable to detect and show these differences.

Managing Complex Applications and Dependencies

Modern IT applications tend to having complex interactions. It's not uncommon for a particular application or service to have dependencies on numerous areas of the IT infrastructure. For example, if an important router or switch fails, users will be unable to access applications regardless of the status of the servers. This makes it important to focus on the end user experience. To ensure high availability and adequate system resources, IT departments must be able to enumerate these complex configurations. Figure 5.5 provides an example.

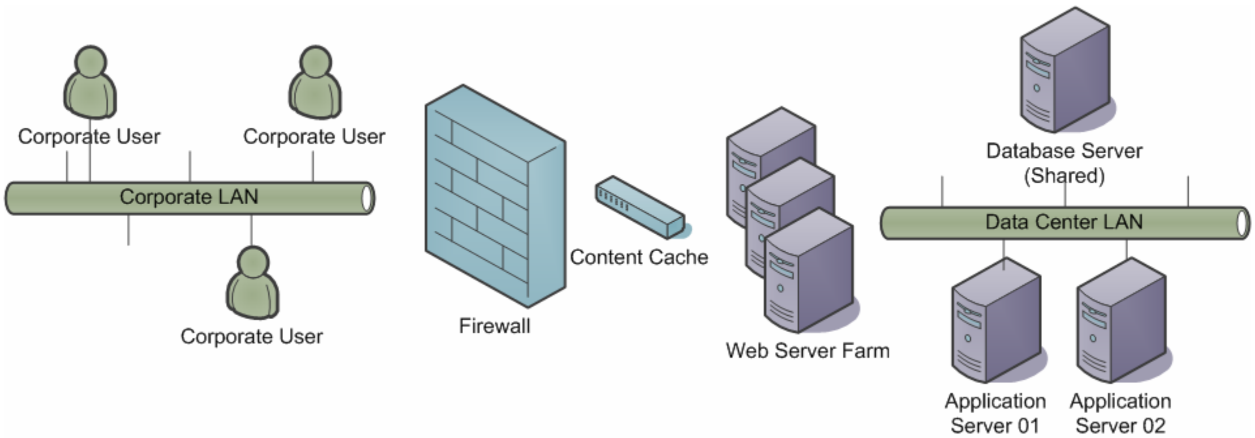

Figure 5.5: Enumerating application dependencies.

In this diagram, there are numerous resources that may be inter-related. If a problem occurs with any one of these resources (such as reduced performance of a database server), numerous applications and services may be affected. It is often difficult to monitor performance consistently when supporting a heterogeneous environment.

Managing Heterogeneous Environments: Best Practices

So far, we've focused on enumerating the many types of platforms that typical IT organizations need to support. We looked at both virtual platforms and physical ones, then listed typical problems that can arise when trying to manage heterogeneous environments. Now, it's time for the good news: There are many ways in which IT organizations can better administer a wide array of platforms. Specifically, we'll look at some examples as they apply to typical management tasks. Figure 5.6 provides an overview of these areas.

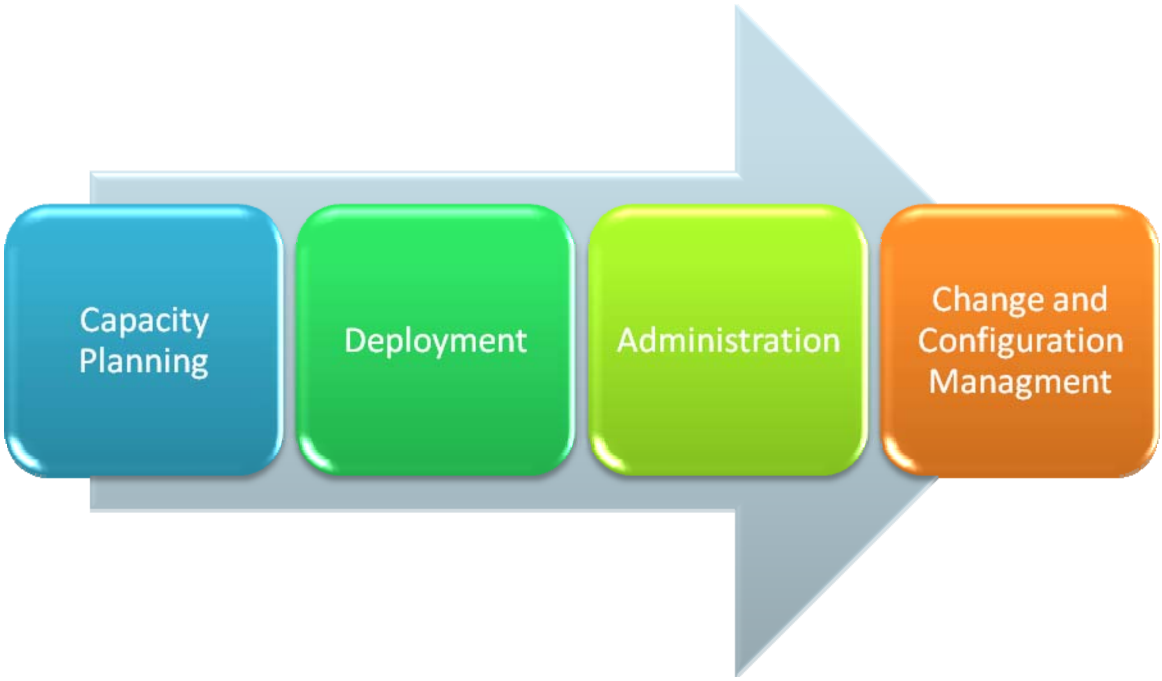

Figure 5.6: Basic areas of management for IT systems.

This section will provide an overview of the general approaches for meeting these challenges. Later chapters will drill down into the details of monitoring and optimizing performance, implementing policies and processes, and automating virtualization management.

Capacity Planning

Many organizations have incorporated virtualization technology into their environments for the purpose of increasing overall resource utilization. Without virtualization, it's common for servers to be running far below their total potential. By using virtualization to add multiple otherwise-incompatible workloads to the same physical computer, organizations can realize significant increases in overall utilization. This approach, however, does provide new challenges. If the overall goal is to maximize the value in hardware and software investments, it is very important to develop plans for determining the best allocation of virtual machines on a group of hosts. This section will look at ways in which environments can optimize the locations of their workloads.

Evaluating Infrastructure Requirements

In the earlier days of IT management, it was fairly common for infrastructure-related departments to work in a reactive fashion. Whenever systems started to slow, they would add the necessary resources to restore performance. As IT departments have moved into a more strategic role, it's important to be able to plan for the deployment of new workloads. There are many infrastructure inter-dependencies between resources for modern applications and services:

- Network capacity—New applications and services often place additional load on network components. Depending on the usage of the particular resource, administrators may need to upgrade Internet connections, inter-site links, and network devices such as routers and switches.
- Shared servers—It's common for certain types of servers, such as Web and database servers, to provide components for many applications. IT staff should be able to predict the reasonable limits for these resources when planning to support new workloads.
- Data center resources—The primary data center factors that must be managed include physical space, power, and cooling. The addition of new workloads will place demands in these areas. In addition, there are often tradeoffs. For example, although virtualization technology can reduce space requirements (by allowing for server consolidation), it's not uncommon for power and cooling requirements to increase accordingly.

Let's look at some details that can help better manage these resources.

Determining Workload Requirements

Perhaps one of the most difficult challenges related to capacity planning is in estimating the true needs of an application. Often, IT staff tends to "play it safe" by overestimating the amount of CPU, memory, disk, and network load an application will require. In many cases, the result is that systems end up with leftover capacity, effectively decreasing hardware utilization. It can also lead to the proliferation of unnecessary server resources throughout the data center. The opposite, however, is at least as bad: If an application or service is not given enough resources, system performance will suffer. Users throughout the organization are likely to be unhappy with the configuration.

The goal in determining workload requirements is to try to estimate resource utilization for an application. There are several main approaches. Table 5.1 provides an overview of the different ways in which performance data can be collected.

Performance Monitoring Approach	Benefits	Drawbacks
Synthetic Benchmarks	• Quick and easy to perform • Can use absolute metrics for resource usage	• Does not necessarily relate to real-world performance requirements
Application Benchmarks/Load-Testing	• Maps to common use cases for applications • Can be automated with load-testing tools	• Can be difficult to set up • May not simulate real-world usage patterns
Historical Usage Data	• Provides real-world statistics based on actual usage patterns	• Not possible for new deployments • Statistics may not be available

Table 5.1: Various methods of measuring workload performance.

Synthetic benchmarks are generally performed using third-party applications. An example might be a disk performance tool that measures disk throughput under varying types of activity (such as sequential and random read operations). The results are metrics that can give an indication of the capabilities of a particular physical or virtual platform. These numbers can also be compared to get an idea of the performance overhead that is caused by virtualization.

In some cases, developers of applications may have created load tests or at least manual sequences of steps that can be used to simulate common usage of the application. IT staff can use this information to generate tests that simulate typical activity and measure the results. For example, an organization might require a particular transaction performed via a Web application to complete within 4 seconds while servicing as many as 50 simultaneous requests. Based on these requirements, IT departments can determine the resources necessary to meet these needs.

Perhaps the best type of performance information is data that is collected in the "real world." Various performance monitoring tools can be used to determine CPU, memory, disk, and network utilization over time. IT decision makers can use this data to determine the best way to deploy the application.

Selecting a Target Platform

In the early days of IT, deployment options were simple. Generally, each application would be given its own server or would be installed on an existing one that had additional capacity. When working in a heterogeneous environment, the decisions aren't always as simple. Systems administrators have numerous options for where to deploy their workloads. Figure 5.7 provides some guidelines for determining which type of option will fit a particular workload.

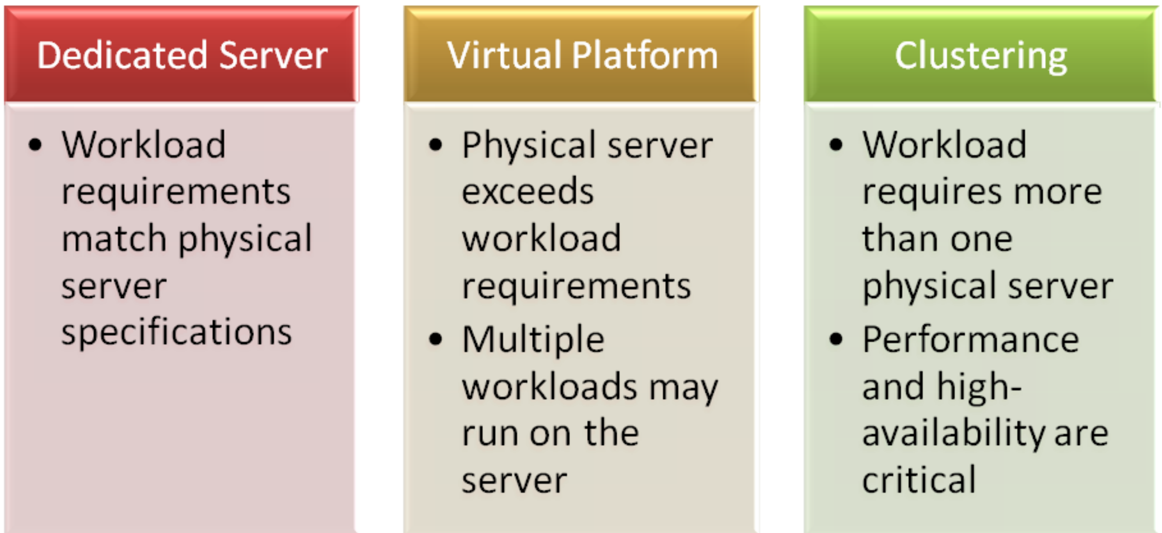

Figure 5.7: Determining target platforms based on resource requirements.

The three cases are broad guidelines and are based on comparing the requirements of an application or service against the available target platforms. In some cases, a particular application might be able to take full advantage of a physical server and utilize all of its resources. That would make the dedicated server options the best choice. More commonly, a workload will be unable to utilize the entire capacity of a computer. In that case, virtual platforms can help increase server utilization. Finally, for some types of production applications, there will be a need to scale out into a cluster configuration. This option will help increase reliability, availability, and overall performance of the workload.

Deploying Workloads

Once IT departments have determined the ideal target platform for a particular workload, it's time to perform the actual deployment. Again, when working in an environment that supports heterogeneous platforms, the methods of deployment may vary significantly. For example, if an application must be installed on only a single or a few servers, it might be easiest to perform the tasks manually. When deployment to many physical servers or virtual machines is required, automation can greatly help. By using scripted methods of distributing workloads, systems administrators can save time, minimize disruption to users, and reduce the chances of errors caused by configuration inconsistencies.

Administering Heterogeneous Environments

A large portion of IT staff members' time is spent on administering the current environment. This includes monitoring and optimizing applications and services that have been deployed on a variety of platforms. Often, the platforms have very different methods of configuration and reporting performance-based information. This heterogeneity can make it difficult for administrators to manage a wide variety of systems effectively. This section will look at some useful practices and features for reducing the administration burden in complex IT environments.

Optimizing Resource Usage

A major benefit of working with virtualization platforms is the ability to quickly and easily reconfigure a computer. Often, a few mouse clicks or commands are all that is required in order to increase memory, add a new network adapter, or add a new virtual hard disk. It's useful to compare this with the process of changing resource configurations on a physical computer. Most hardware modifications will require the server to be powered down. Adding resources such as CPUs and memory requires direct physical access to the hardware itself. Data center administrators often have to un-rack the server, make the modifications, and then re-rack it. The process can take a significant amount of time and effort (especially if it must be done during a scheduled downtime window).

Given that it's easy to reconfigure virtual machine characteristics, it's important to assign the appropriate amount of resources to a workload over time. The goal is to ensure that systems are being run at the maximum practical level of utilization.

Earlier, we talked about how administrators can estimate resource requirements prior to deployment. It's inevitable that changing usage patterns will require modifications to this configuration. For example, changes to the organizational structure might add many more users to an application than was originally estimated. Or, the opposite may be true: A service may be no longer needed for production use. These situations call for the reconfiguration of a virtual machine. Perhaps additional memory should be added to a virtual machine or it might benefit from additional disk space. Although these operations can be performed manually, it quickly becomes difficult to manage large heterogeneous environments.

Automated virtualization management solutions can track performance and resource usage in real-time and make the required changes automatically. If, for example, a particular virtual machine has been using far less than its allotted amount of memory for the past several days, its memory limits can be decreased. The remaining physical memory can then be provided to another workload that might be experiencing excessive swapping due to low memory.

Movement of Workloads

Although changing the configuration of a virtual machine can often meet the needs of many performance-related issues, it is sometimes necessary to move an application or service to another computer. In some cases, a virtual machine might have exceeded the physical resources that are available on a particular host computer. In other situations, two workloads might be competing for the same resources. In these cases, it's often best to move the virtual machine to another host computer.

When performed manually, there are numerous steps that are generally required in order to move a virtual machine between systems. Figure 5.8 provides a list of the process that is usually involved in moving a virtual machine from one host to another.

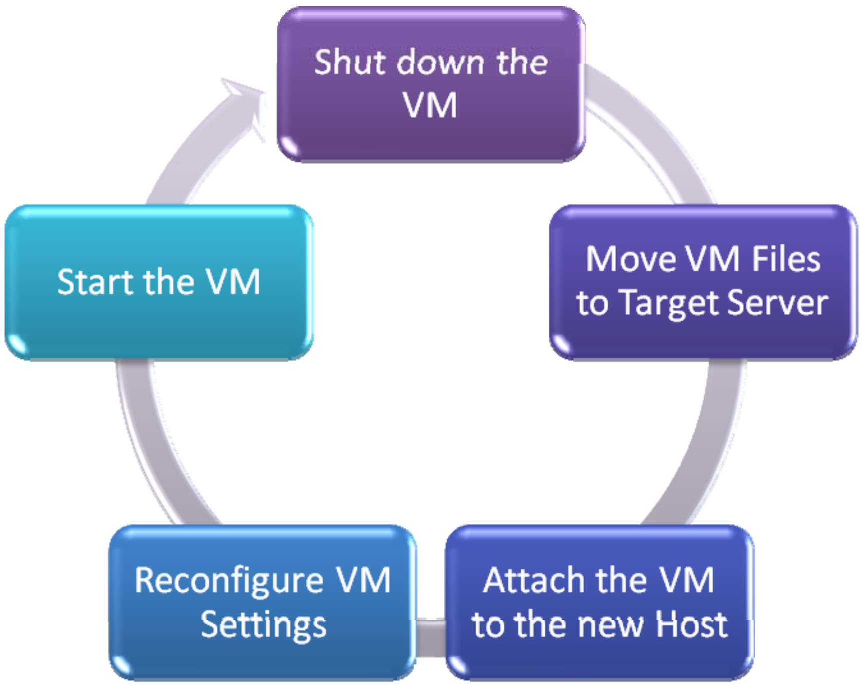

Figure 5.8: Steps involved in moving a virtual machine from one host to another.

Often, the amount of manual labor that is required to move a virtual machine limits the number of migrations that are performed. Ultimately, this can lead to some physical servers being over-utilized while others are barely using a fraction of their potential. Automated management solutions can help simplify the movement of virtual machines between host servers, thereby simplifying the process. The end result is increased resource utilization across the environment.

Unifying Physical and Virtual Management

The typical characteristics of a workload remain fairly uniform regardless of the platform to which it is deployed. Common components include:

- The application or service itself
- Required supporting services
- OS settings
- Hardware (or virtual hardware) configuration

As an example, consider a Web server that is designed to provide access to a Customer Relationship Management (CRM) system. Whether the workload is running on a physical server or within a virtual machine, the same techniques will be used to measure performance, make configuration changes, verify security settings, and perform general administration tasks. Automated management solutions that provide a unified method of administering these systems—whether they are hosting on virtual or physical platforms—can be very helpful to systems administrators. The benefit of providing consistent management methods include reduced training time, consistent sets of practices, and efficient management of changes.

Centralized Management

A major source of frustration for systems administrators is the need to access numerous tools to perform administrative tasks. It's not uncommon for a single individual to have to manage hardware, applications, services, and network devices in a single day. In simple environments, it's certainly manageable to perform these tasks manually. However, in larger environments—especially those that support heterogeneous platforms—the tasks can take significant time and effort.

One potential solution to this problem is the use of a centralized management application. Generally, this type of product is designed to provide a single console from which all types of platforms can be managed. Figure 5.9 provides a conceptual overview.

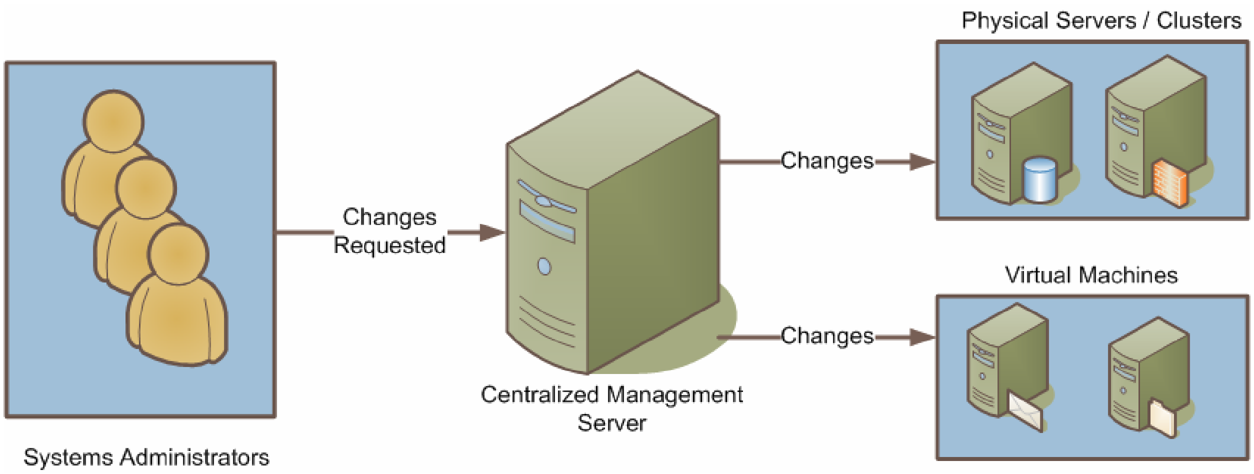

Figure 5.9: Using centralized management in a heterogeneous environment.

In this diagram, systems administrators and other IT staff connect to an administration application to monitor systems and make configuration changes. It is the responsibility of the application itself to contact different platforms to obtain information and commit modifications. This layer of abstraction allows the management console to provide a unified view of important workloads, regardless of whether they're running on physical or virtual platforms. The result is the ability to consistently and efficiently administer applications and services throughout the environment.

Broad Platform Support

The rate of change in IT tools and solutions can simultaneously be the best and worst parts of administering heterogeneous environments. On the positive side, the availability of new platforms and versions often means that systems will be able to run more efficiently. There are often compelling reasons for wanting features of new products, as they are sure to pay off in the long run. But what about managing even more heterogeneous platforms? Supporting several different virtualization solutions can be difficult enough, and adding another thing to the mix might be more work than the IT staff can handle.

When evaluating an automated management solution, it is important to consider the types of platforms that are supported. One important consideration is the current list of supported platforms. Ideally, the list will include all the current platforms that an organization is using as well as any that are planned for the future. In general, the more platforms, the better.

However, given that change is likely, it's quite possible that a new platform will emerge in the near future. Although no application vendor can predict the future with any level of certainty, the ability to easily expand an automated management solution can be a huge benefit. In some cases, vendors might be able to provide a simple software update or a management module that provides the ability to fully manage a new platform. Ideally, the solution will be able to utilize the same user interface and other features without requiring retraining of the IT staff. This type of extensibility is critical in the world of platform management, as none of us can be certain which platforms we'll be supporting in the future.

Automated Discovery

One of the first—and, often, most difficult—tasks related to managing an IT environment is keeping track of everything that's "out there." This is especially important when dealing with heterogeneous environments that include virtualization technology. Virtual machines can be deployed in a matter of minutes, and it's often difficult for IT staff to keep track of where and when these virtual machines have been deployed. A potential approach to managing these systems is a manual one: Systems administrators can periodically log on to physical computers and inspect what is running. However, even with relatively few physical servers, this method becomes error-prone and difficult to scale.

Management solutions that offer automated discovery methods can make this process significantly easier. The discovery process often begins by scanning the network for server computers and network devices such as routers, switches, and firewalls. In the world of physical assets, this approach works very well.

However, virtualization presents an additional challenge: It's possible for virtual machines to be configured without network access, or for them to be powered-off at a specific point-in-time. For these reasons, management solutions should have the ability to query physical virtualization host servers and to be able to identify any virtual machines that are defined on those computers. Additionally, the process should be run regularly to detect any new deployments and to determine when workloads have been moved to other systems. Using these features, systems administrators can keep track of even the most fluid IT environments.

Change and Configuration Management

In addition to standard systems administration, it often becomes important to make changes to existing workloads. In some cases, modifications are needed for business reasons. In other cases, changes to the technical landscape necessitate upgrading or moving an application or service. Regardless of the reason, it's up to IT staff to make sure that changes are made carefully and that systems are not overlooked. This section will look at ways in which changes can be managed in heterogeneous environments.

Developing Virtualization Standards

Most virtualization platforms provide a wide array of configuration options for each new virtual machine that is created. Common options include:

- Amount of allocated physical memory (RAM)
- Number of virtual network adapters
- Network connection details
- Virtual hard disk types (fixed-size or dynamically expanding options)
- Hard disk controller types (usually IDE or SCSI)
- Removable media devices such as floppy drives
- USB-based devices
- Additional connections, such as serial and parallel ports

The primary downside of this issue is that the number of configuration variations can quickly become very large. A recommended practice is for IT departments to define several standard virtual machine configurations based on users' typical needs. For example, a "Standard Windows Server" virtual machine configuration might have 1024MB of RAM, two virtual network interfaces, and two virtual hard disks. Common configuration variations, such as the use of additional memory or devices, could form another standard. Overall, the goal is to enforce consistency in virtual hardware configurations.

Additionally, IT organizations should develop a consistent naming convention for their virtual workloads. In the world of physical servers, computers are generally named based on their rack locations. Virtual workloads can be easily moved between systems, so this convention is not as useful. Naming should involve using a description that includes details and the workload type (for example, a Web or file server), the virtualization platform (for example, VMware ESX Server or Microsoft Virtual Server), the guest OS (such as RedHat Enterprise Linux), and some form of unique identifier that can be used to distinguish it from other virtual machines with a similar configuration.

Support for a Virtual Machine Library

Earlier, when discussing challenges related to managing virtual environments, we saw how the ability to quickly copy or move a virtual machine can make administration of the environment more challenging. Often, a single user can create a new virtual environment in a matter of minutes, and with limited technical knowledge. Very quickly, IT departments can find themselves supporting dozens of configurations, many of which were never reviewed or approved. Additionally, many of these virtual systems will lack the benefits of having a thorough configuration review by IT staff prior to deployment.

One method of limiting configuration variations is to require end users and systems administrators to rely on a virtual machine library. This library is a collection of base virtual machine images that are created and certified by IT departments to be supported configurations. Figure 5.10 provides some examples of configurations that may be used.

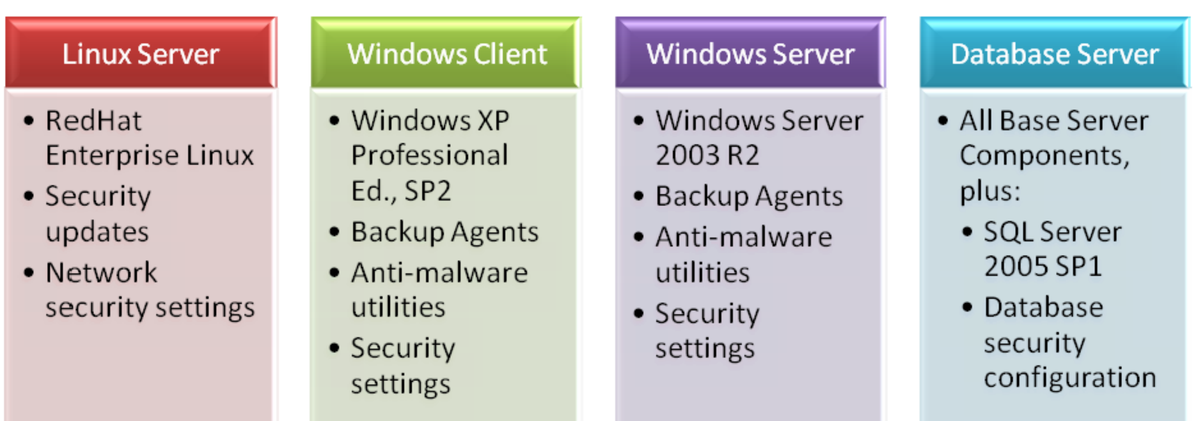

Figure 5.10: Examples of images in a virtual machine library.

Once they have been created and configured, virtual machine libraries are often made available on file servers. IT departments can benefit from additional controls, such as reviewing all configurations prior to their deployment in a production environment.

Unfortunately, it's not realistic for IT organizations to expect to meet 100% of users' needs with these base images. In cases in which other OSs or applications are required, IT departments should make their support policies clear and should limit the potential impact of the deployment. For example, if a group needs to run a seldom-used set of OSs for testing software compatibility, IT departments may set up an isolated network to prevent them from affecting production operations.

Maintaining Security

Keeping OSs and applications up to date is an important concern related to maintaining security. Many vendors have created automatic patch deployment tools and methods to help organizations with specific platforms. However, the problem can be complicated when running numerous different types of platforms. For example, on virtualized computers, the host OS and guest OSs must be kept updated.

Automated management tools can simplify the process by providing a consistent method of patch deployment, regardless of the platform. An added benefit is the ability to quickly scan a large number of computers over the network and to identify any machines (whether virtual or physical) that are lacking important updates. These methods can help ensure that all production workloads are adequately protected.

Role-Based Administration

In an ideal world, one or a few systems administrators would be able to manage all the tasks that are required to keep an entire data center running optimally. Realistically, the various roles and responsibilities are often delegated to numerous individuals—each of whom will have their own set of permissions and job responsibilities. Automated platform managements should provide for the ability to define roles for each user of the system. For example, some systems administrators might be able to create new virtual workloads and deploy them to specific servers. Others might have the ability to install updates and make configuration changes. Roles can even extend outside of the IT department in a "self-service" model. For example, users might be able to carry out the creation of a new virtual machine based on a library image without requiring any assistance from the IT department. Overall, the goal with role-based administration is to ensure that security and accountability are retained while making administration as simple as possible.

Integration with Other Management Tools

Many organizations already have investments in enterprise management tools. Many of these tools follow standard best practices for collecting and recording configuration information. One example is the use of a Configuration Management Database (CMDB)—a centralized database that stores information about all the hardware, software, and network devices that an IT environment supports. The use of a CMDB can help ensure that information is stored centrally and that it is readily available for running reports.

Automated virtual platform management tools will ideally have the ability to integrate with an organization's existing change and configuration management implementations. Ideally, the solution will plug in to the architecture seamlessly, while adding support for many types of virtualized workloads. The end result is unified management of a complex environment.

Enforcing Policies and Processes

To maintain consistency and order in their production environments, many IT organizations have developed policies and processes for determining how and when changes are made. The purpose of these guidelines is to ensure that tasks are carried out in an optimal way. Figure 5.11 provides an example of a process that might be followed to address a request for the deployment of a new application or service.

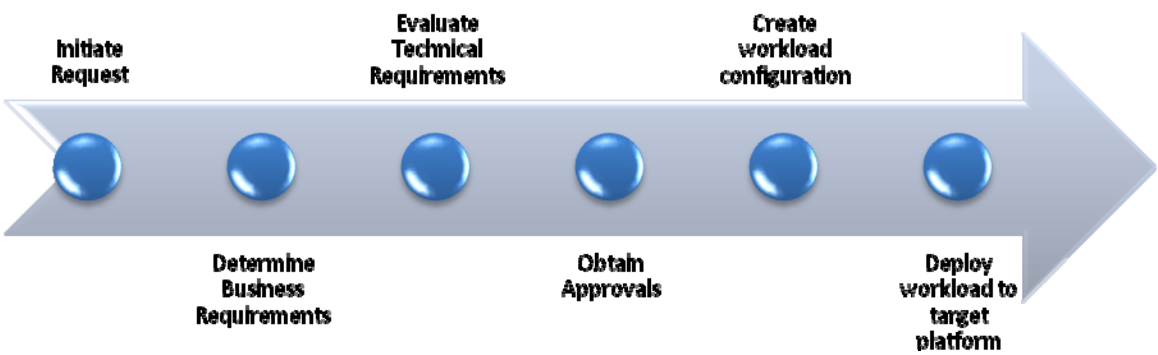

Figure 5.11: An example of a workload deployment process.

This process is relatively simple, and in smaller organizations, it might be accomplished in a single meeting. Larger organizations will often require input from numerous individuals before the deployment is complete. Regardless of the overall scale, however, the goal is to ensure that there is a consistent set of steps that must be carried out to perform the deployment of a new workload into the environment.

Summary

This chapter covered important details related to managing heterogeneous platforms in a typical IT environment. The chapter began by enumerating the platforms themselves. First, we looked at virtual platforms, including application virtualization, desktop virtualization, OS virtualization, and server virtualization. We then looked at physical systems: client computers, rack-mounted servers, blade servers, and clustering. Most IT organizations will be required to support most (if not all) of these platforms.

Next, we looked at the types of challenges that systems administrators face when trying to manage heterogeneous environments. Issues include virtual machine sprawl, having to use numerous management tools, dealing with resource allocate issues, and managing complex applications that are deployed on virtual and physical servers. All of this helped to set the stage for best practices for managing heterogeneous environments. We looked at ways in which management tools can be used to simplify and automate the process of administering workloads, regardless of the specific platforms on which they're running. The specific steps included capacity planning, deployment, administration, and change and configuration management.

Upcoming chapters will revisit several of these topics in more depth and explore specific details of how heterogeneous virtual platforms can be better managed in production IT environments.