Part II

Planning

Chapter 4

Business Cases for Server Virtualization

The first step in planning to deploy server virtualization technologies is to understand how it can be applied to business objectives. This chapter describes the key areas of information technology to which server virtualization can be effectively applied and how it relates to common business goals. It also shows how common short-term and long-term problems can be addressed using server virtualization technology and where it should not be used.

Solving Business Problems

Aside from the pure novelty of running more than one operating system on a single computer at the same time, server virtualization technology was designed to be used to help solve business problems. Virtualization can be used in a very straight-forward manner much like physical servers and when used in this way can help reduce physical server hardware costs by consolidating multiple servers onto a single server. This usage is commonly referred to as server consolidation. Virtualization provides new features that are either unavailable or unobtainable on physical server hardware. One of the most profound benefits of virtualization is the abstraction between virtual machines and the physical hardware, thereby making virtual machines portable. In this way, virtualization simplifies and enhances legacy server and application support, while providing new opportunities in disaster recovery and high-availability scenarios. Because virtual machines are much easier to deploy than physical servers and because the software nature of virtualization technology lends itself well to automation of provisioning, on-demand and adaptive computing are more easily realized.

Server Consolidation

Many organizations today have several servers, each performing a dedicated task. These servers include file servers, print servers, e-mail servers, Web servers, database servers, and other application servers. It is not uncommon to have applications that require or strongly recommend a dedicated server, usually because the software does not particularly coexist well with other applications installed within the same operating system. In these situations, organizations will purchase small dedicated servers to host these applications, many of which have a low rate of utilization because the application only has a small amount of users or because it may only be used infrequently, such as once a month. The problem with this scenario is that the organization must incur a capital expenditure of $6,000 or more and that the processor, memory, and disk storage of the new dedicated server are highly underutilized and essentially wasted. Over time, the data center may support many small, highly underutilized servers hosting dedicated applications, also incurring the data center hosting costs (ping, power, and pipe) for each server.

Server virtualization can help organizations save money in the datacenter through server consolidation. Many, if not all, of the organization's highly underutilized application servers may be moved from physical hardware into virtual machines. Applications requiring a dedicated server still meet that requirement because a virtual machine is a dedicated server with its own, fully isolated virtual hardware and operating system. Many of these virtual machines may be hosted on a single physical server by means of the virtualization platform. As new dedicated application servers are needed, instead of purchasing new physical servers, virtual machines may be created on existing virtualization host servers. The density of virtual machines, which may be hosted on a single virtualization host server, depends on several factors including the number of processors, the amount of memory, and disk storage space on the physical host server, the virtualization platform (installed on the host server), and the amount of resources consumed by the virtual machines. It is not uncommon to have 10:1 or higher server consolidation ratios for highly underutilized servers.

Legacy Server and Application Support

A common pain point of many information technology organizations is that, over time, legacy hardware and software must be maintained. The term legacy commonly refers to either outdated and possibly unsupported computing systems or systems comprised of components of a previous version. Replacing legacy systems with new systems is always a challenge, both technically and in business terms. Legacy systems exist because of many reasons. These may include a very high cost to completely replace the system due to a large investment in capital expenditure, training, and customization of the system. Legacy systems may also be sensitive, mission critical resources that cannot easily be replaced. Sometimes legacy systems must be maintained because they cannot be replaced because it was purchased from a third-party company no longer in business or from one that no longer supports the system, or possibly because of a lack of expertise in the system's internals. In some cases, it costs less to maintain legacy systems than it does to replace them, but over time legacy systems typically become more unusable because they often are not able to scale up because of costs or technology limitations.

For example, an organization may have invested in a custom software application designed to run under the Microsoft Windows NT 3.51 Server operating system many years ago. The application in this example was written in such a way that it will not work with any other version of the Windows operating system. The application is becoming a productivity bottleneck because its utilization has grown over the years while running on the same hardware. The organization wishes to upgrade the server hardware on which the application and operating system are installed upon, but cannot because the newer hardware available today does not have the appropriate support for the older operating system. There are no available chipset and storage drivers that will work with the older operating system while allowing it to use top-of-the-line processors, memory, and disks to solve the legacy application's ability to scale up. Additionally, because the application was custom-written, there are no newer versions of the application available and the costs to switch to a different application that provides like functionality is cost prohibitive.

Server virtualization can solve legacy server and application support issues. Migrating the legacy servers to virtual machines inherently abstracts the physical hardware from the legacy software, including the operating system and the applications, allowing the legacy hardware to be discarded or reused elsewhere as needed. Because the legacy software now uses virtualized hardware, it can be moved to any host server as necessary, making the legacy server portable. Aside from its newly gained portability, the legacy server migrated to a virtual machine can be hosted on and use any hardware that is supported by the virtualization platform and the host server.

Continuing the previous example, the organization decides to migrate the legacy server to a virtual machine. They choose a virtualization platform that allows them to host the migrated server on modern, high-end server equipment, which addresses the bottleneck issues while adding the portability necessary to move the server to better hardware in the future, if the need arises.

Additional savings in terms of hosting can be realized by migrating legacy servers to virtual servers in order to decommission older, inefficient hardware and by increasing overall server utilization.

Disaster Recovery

Disaster recovery is a term used to describe the strategy and processes used by an information technology organization to ensure that in the event of a disaster that damages or destroys the organization's computing infrastructure, a minimal recovery of the computing infrastructure and data can be accomplished in a reasonable amount of time to provide business continuance. Disaster recovery strategies are almost always an afterthought in most organizations, probably because it is like paying for insurance. When an event occurs, however, it is the disaster recovery strategy that can save a company from going out of business, thus providing business continuance. Although disaster recovery strategies encompass a wide range of considerations, one of the most common activities is data backup.

Backing up the data of an organization is critically important to help not only with disaster recovery, but also to help to recover from day to day mistakes, such as when a user inadvertently deletes an important document, or to help protect data lost from a malicious intruder. There are many theories and methods of performing data backup. Most commonly, only an application's data is backed up instead of the application itself or the operating system. This is typically done to save space within the backup system (and to reduce the cost of the backup system) because the application's code and the operating system itself changes rarely. It is also traditionally more difficult to restore an operating system and applications from backups versus reinstalling them and then applying data afterwards. Because of these methods, the restoration of systems becomes a much harder, longer task. Although complete system restoration is rare (hopefully), the amount of time it takes to restore systems is critical to business continuance.

Virtualization technology can help improve disaster recovery strategies in many ways. Since virtual machines, their guest operating systems, all installed applications, and data reside within one or more files on the host server, it is simple to backup an entire virtual machine. To backup the virtual machine, make a copy of the virtual machine's configuration files and its virtual disk files. Backing up entire virtual machines usually does require more backup storage space because of the size of the virtual disk files, but the benefits are worthwhile. In the event of a disaster, the entire virtual machine can be restored in the time it takes to recopy the virtual machine's configuration and disk files. Installation and configuration of neither operating systems nor applications is necessary. In the same amount of time it takes to rebuild a single standard Windows server, configure it, apply patches and updates, install applications, and restore data, many virtual machines can be restored, possibly 10, 20, 30, or more.

Another aspect of using virtualization in regards to disaster recovery is that when a disaster occurs and a recovery must take place, the proper hardware components must be obtained on which to recover the systems. Traditionally, older hardware is stored offsite to be used for recovery and only the core, mission critical servers can be restored because of hardware limitations. Any secondary systems cannot be restored until more hardware is acquired. Using virtualization, a few physical servers able to support the virtualization platform being used are required for recovery. The most critical virtual machines can easily be restored onto any of the virtualization platform-compliant hardware and shortly after that, other secondary systems may also be restored onto the same servers utilizing the unused hardware resources (much like server consolidation).

Virtualization can also help aid in disaster recovery even if not all of an organization's servers are virtualized. For instance, high-performance, highly utilized file and mail servers may not be virtualized in production. Other servers within the organization are consolidated using server virtualization. As part of the organization's disaster recovery strategy, there are virtual machines already created but not used that replicate the functionality of the file and mail servers. When the time comes to recover these servers, the virtual versions provide very quick service restoration times and will usually be adequate until the proper hardware can be acquired.

High Availability

Although disaster recovery strategies address business continuance in the event of a catastrophic disaster, they do not address the needs of fault tolerance and system robustness. High-availability strategies provide business continuance through the use of fault-tolerant systems, usually by implementing redundant system components such as RAID storage systems or clustered servers. Most highly available systems must be available more than 99 percent of the time, meaning that the systems must be able to recover from a number of different component failures. One of the most common methods of providing a highly available system is through the use of clustered servers. In a server cluster, there are at least two servers that have the same applications and configuration and share access to the same data. When the active server in the cluster experiences a fault (such as a failed network adapter or hard drive failure), one of the other servers in the cluster take over and become the active server in order to provide nonstop services. One of the issues associated with server clusters is the additional cost of server hardware that is mostly unused while it is waiting for a failure to occur.

Server virtualization can provide an inexpensive method of implementing server clusters by using virtual machines in a cluster. Virtual machines can be clustered with other virtual machines or other physical servers. Using virtual machines in server clusters works well with server consolidation. While the primary server in the cluster is active, the secondary server, a virtual machine, will usually consume very few resources, allowing it to be placed on a host server along with other servers. If the primary server in a server cluster is also a virtual machine, it should be placed on a separate physical host server in order to maximize the cluster's uptime.

Adaptive Computing

Another area where server virtualization provides business value is in adaptive computing. Adaptive computing consists of server systems that have the ability to autonomously reconfigure themselves to address changing requirements. Adaptive computing is also referred to as autonomous computing, grid computing, on-demand computing, or utility computing. On-demand computing as referred to in this book is different from adaptive computing.

Server virtualization can work well with adaptive computing initiatives because of the ease of virtual machine provisioning. For example, consider a bank of Web application servers in a load balanced cluster. The Web application utilization rises and the overall performance of the application decreases. The system then allocates resources on a virtualization host server on which it creates two additional virtual machines using the same Web application server image. Once the two virtual machines have been created they are booted up and added dynamically to the existing cluster. The two additional servers help spread the application's workload over more computing resources, thus increasing the overall application performance. When the application's utilization falls off, the two additional servers are no longer needed and they are powered off and deleted.

This type of adaptive computing can be applied to many applications that share a common set of virtualization host server resources on which to dynamically create virtual machines. In addition to dynamically responding to needs, adaptive computing systems can have capacity scheduled in order to help optimize computing resource utilization. For instance, during the week an application may have five virtual machines on which to perform its work, but over the weekend, three of the virtual machines may be scheduled to be reconfigured to work with a different application to help with back-end processing.

Adaptive computing scenarios can be achieved with physical hardware, including traditional server and blade servers, but typically at an increased cost and increased level of complexity as compared to using server virtualization.

On-Demand Computing

Virtualization technology easily facilitates on-demand computing systems that can quickly allocate one or more servers to a group of users for a short time and reclaim and reuse those resources once the users' time has expired or if they have released the resources. This type of system is commonly used to provide virtual labs used by system administrators and developers for testing software configurations. It is also used to provide virtual training labs within a learning management system, whether it is a physical class or an e-learning system accessed over the Internet.

Once again, the ease of provisioning virtual machines supports on-demand systems very well, but it is the virtual machine's ability to discard changes made to their virtual hard disk drives, resetting them to a pristine state, where the most value is gained. Building these types of systems without using virtualization is very difficult and expensive.

Limitations of Server Virtualization

Virtualization technology opens the data center to new possibilities that may provide cost savings and new types of functionality, but virtualization does have its limits. Not every server or application is well suited to be run in a virtual machine. Some applications are highly performance-sensitive, such as databases, data warehousing applications, business intelligence, reporting, and many others. These applications usually require multiple processors and massive amounts of memory. At the present time, server virtualization platforms support virtual machines with either a single processor or at most dual processors and up to 3.6GB of memory. In the near future, these limitations will likely be overcome, although for performance-sensitive applications, it may still not be enough to justify running them as virtual machines. This is because virtual machines are not only sharing resources with other virtual machines on the same host server, but because the virtualization platform itself incurs a small amount of overhead that can negatively impact performance-sensitive applications.

Virtualization technology does not support graphics-intensive applications very well at this time. Graphics-intensive applications and games usually require the use of high-performance video cards. In the virtual machine, the video card is a virtual device implemented in software and at this time does not support advanced graphics acceleration features. Even if the virtual video card device did have support for advanced graphics acceleration, it would invariably place more overhead on the system's physical processors.

Much like the problems with graphics-intensive applications, applications requiring specialized hardware cannot be used within a virtual machine. Specialized hardware refers to any nonstandard hardware device, commonly PCI cards. At the present time, virtualization platforms are generally closed systems and new types of virtual hardware cannot be created by third-parties. In the future, this limitation may be overcome, at which time third-parties may have the ability to write their own virtual hardware devices in order to connect specialized physical hardware devices to virtual machines.

Summary

Server virtualization can be used in many ways to help solve business problems. Server consolidation can reduce data center costs while enabling the virtualized servers to be backed up in an easier manner helping with disaster recovery. The lifetime of legacy servers and applications can be significantly extended and the performance increased by moving them to virtual machines. Disaster recovery strategies can be enhanced through the use of server virtualization, which can dramatically reduce the time and complexity required to restore vital systems in the event of a catastrophic event, improving the chances of business continuance. Server virtualization can also help reduce the costs involved in creating highly-available systems, also aiding in business continuance through increased fault tolerance. In addition to cost reduction, server virtualization opens up new opportunities by providing the capabilities to easily create adaptive and on-demand computing systems increasing an organization's productivity. Server virtualization is not the answer to all business problems, however, and some systems do not lend themselves well to be run within virtual machines, including performance-sensitive applications, graphics-intensive applications, and applications requiring specialized hardware. It is important to take these limits into consideration when planning for server consolidation projects, legacy server and application support, disaster recovery, and high availability solutions. All in all, virtualization can save the organization thousands to millions of dollars through more efficient and effective use of hardware.