Realtime publishers

The Essentials Series:
Virtualization and Disaster Recovery

Data Protection in a Virtualized Environment

sponsored by

by J. Peter Bruzzese

Copyright Statement

Data Protection in a Virtualized Environment

In a traditional network environment, the hardware is the heart of operations. But times are changing, and although hardware is still an indispensable part of the operations in any IT center, software technologies are expanding into new territories. These technologies aim to provide the same—or better—level of strength found in a hardware solution while decreasing the number of hardware devices you need in your business environment. One of the most popular technologies in this regard is virtualization.

An Overview of Virtualization: Enhanced Protection with Less Hardware

Virtualization is not a new concept—it has existed for decades—but what is new about it is the rapid rate of adoption in IT departments of all sizes and shapes. Once used mainly in testing and development, virtualization is now a key player in production environments. In the event you haven't had a chance to work with virtualization software, let's review the overall concept.

The concept of virtualization is simple—you run a piece of software (a virtual machine) on a PC as if this virtual machine is just an ordinary software application. In the virtual machine, you can install the operating system (OS) of your choice and on top of this OS, install any applications you need. This way, on one hardware configuration, you can have many virtual PCs, each of them running an OS and software applications of its own, independently from the OS and the applications you are running on the main (host) computer. Keep in mind, however, that the layer running above the hardware is standing in-between your virtual systems and the bare metal, so to speak. The thinner that layer can be the better.

What stands between the hardware and the guest OS is called a *hypervisor*. Typically on standalone systems, the hypervisor runs off the conventional OS that you might be using, whereas a hypervisor that runs directly on the bare metal or as close to it as possible is a more server-oriented virtual machine (VM) architecture that you would look to when considering virtualization and data protection.

Note

Although multiple virtual machines running on one physical machine illustrate what server virtualization is, there is another equally important type of virtualization that is also used in IT environments—storage virtualization. Server virtualization uses the hardware resources of one computer to run multiple OSs/servers/applications; storage virtualization deals with storage devices only (for instance, disk arrays). Similar to server virtualization, which abstracts logical servers from the physical hardware infrastructure they run on, storage virtualization abstracts storage from the separate drives where data is physically dispersed. Briefly, data can be written over multiple physical disks, but from the administration console, all these disks appear as one logical entry. The term *storage virtualization* might not be as mainstream as its server counterpart; there is hardly an IT person who has not heard of RAID, SAN, or NAS, which are examples of storage virtualization.

Typically, when one says "disaster recovery," RAID technologies are the first to come to mind. Although storage virtualization technology is a more common direction people look to when attempting to protect their data and system availability, server virtualization is becoming so reliable and cost effective that disasters are no longer mitigated through traditional backup/restore methods. In fact, using virtualization technology, a company may be able to achieve 100% availability with complete data protection.

Cost Savings of a Mirrored Site (Fewer Systems and Personnel)

Server virtualization, real-time data replication, and failover technology are capabilities that provide a high level of protection while keeping costs at a minimum. As already explained, server virtualization allows you to maximize performance by allowing fewer virtualized servers to do the work of multiple physical servers, yet at the same time cut costs because—when compared with non-virtualized environments—server virtualization allows for fewer machines to perform the same amount of work.

Where virtualization really shines for data protection and disaster recovery is in the ability to eliminate the typical hardware equality required in physical server redundancy and recovery solutions. The pieces of hardware become less critical in a virtualized environment. By moving to a virtual model with a hardware-agnostic formation, you can restore systems without a concern for identical hardware or even near-identical hardware. This provides a quicker and far more efficient disaster recovery solution.

Host servers must be powerful machines (in terms of CPU and RAM mainly). Although the minimum requirements of most server virtualization solutions are moderate, optimal performance can't be achieved on a weak machine. In any case, adding more RAM or upgrading to a faster CPU is much cheaper than purchasing a separate brand-new machine—one of the points that make server virtualization a cost-effective solution.

One of the biggest obstacles to choosing an effective protection solution is cost. If a company's fiscal needs were not an issue, there are many advanced hardware and software solutions that provide a really robust data protection mechanism. However, the reality is that very few, if any, IT departments can afford to invest millions in such a robust system. Of course, in those cases, where data protection is more than vital (for example, military systems), cost comes second to reliability.

There are many robust solutions that might not necessarily be military strength but are vigorous enough to protect critical data. Very often, the optimal solution is a high level of data protection at an affordable price; these solutions are what a typical IT department needs.

In addition to cutting hardware costs, virtualization allows for a reduction in the number of required personnel, thus providing for wage-related savings. Virtual environments can be centralized and managed remotely, so there is need for fewer personnel working on the administration of systems, and your workforce can now be used for other tasks. Certainly, in a challenging economy with possible shrinking budgets being offered to IT departments, the ability to decrease personnel resources is a motivator.

Facilitation of Testing

As I have already mentioned, before going mainstream in production environments, virtualization was used mainly for testing. The uses of virtualization for testing have by no means decreased and even companies that don't use virtualization in a production environment use it daily for testing.

The advantages of virtualization for testing are obvious. Having your data virtualized allows for more frequent testing than with physical servers. You just copy the testing setup from another machine or restore it from a backup, and you are ready to go. There is no need to install and configure applications or OSs—you simply copy the file of the virtual machine. In addition, while physical server testing can oftentimes be painful and frustrating, virtualization allows servers to be shut down and turned on easily, even allowing for remote server failure simulation without concern over physical access to the system that you normally prepare for when testing remote failures on physical systems.

However, you should be aware of the potential downsides of this solution. With virtualization for testing, one of the drawbacks is that it is not an impossibility (though certainly it is not the norm) to observe differences in the behavior of applications when run on a virtualized platform than when run on standalone machines. In some cases, the differences might be critical and, in even rarer cases, it is possible that there are applications that cannot (or shouldn't) be run in a virtualized environment. Fortunately, such cases are infrequent; thus, typically, it is appropriate to use virtualization for testing.

The Rules Regarding Traditional Data Protection Are Not Deprecated

When discussing data protection in a virtualized environment, the perception might be one-sided in thinking that virtualization is useful only for remote site servers supporting replication and failover. It's true that virtualization provides for amazing recoverability through WAN disaster recovery sites, but it's important to remember that you must have a backup system designed into your production environment to ensure the protection of both physical and virtual servers.

Data protection starts at the physical level. If there are no clearly defined rules about how you protect data physically, even the most sophisticated methods are easily compromised. The physical aspect of data protection includes access, backups, and protection against natural disasters, and it is the basis of your data protection policy.

You must have clear rules regarding who and under what circumstances data can be accessed. Unauthorized access makes any other data protection methods useless. You should also regularly back up your data (weekly, daily, or even real-time) and keep the backups in a safe place. Natural disasters don't happen on a daily basis but when they do happen, the result is devastating. Therefore, if you want to make sure that your data is protected against physical destruction, always make reasonable measures to protect the premises against flood, fire, earthquakes, and so on, and above all, always keep an up-to-date backup at a secure remote location.

To fully enjoy the benefits of virtualization for data protection, you need to apply certain measures to ensure that the data is protected from risks. Virtualization is not a panacea, and if you neglect basic rules in regard to data protection in general and in a virtualized environment in particular, don't count on virtualization to save your data. When data is dispersed over a network, no matter whether it is a LAN or a WAN, and this network is not guarded properly, there is no effective data protection mechanism in place. A network is as secure as its weakest part, so if there are gaps in the security policies (that is, roles and privileges are not properly defined and enforced), the integrity of the whole system is under question. Generally, the least privilege rule is a great security policy, so you're better being restrictive now rather than sorry later.

Move Data Quickly and Cost Effectively Across the WAN

When disaster strikes, the last thing you have is time to think what to evacuate first. Usually, even if there is a detailed disaster recovery plan (and such a plan should be part of any data protection policy in every IT department) and you know what you must do, time is a critical asset and you don't have long hours at your disposal.

Thus, you need data protection solutions that have the ability to move data quickly and cost effectively across the WAN. If you can count on greater utilization of the links involved in the WAN portion of your protection plan, you have much more flexibility to plan properly. Combining virtualization with a solid WAN connection to a disaster recovery site changes the recovery game dramatically. You can replicate both physical and virtualized servers over to virtual servers at the disaster recovery location both quickly and cost effectively.

Virtualized servers are a boon in terms of speed of data evacuation. Because virtualized servers are in fact just software files that you can copy from one machine to another in no time at all—or can be copied automatically through specialized solutions that provide for just that type of scenario—your entire set of operations (servers, applications, and files) can be redundant in an alternate virtualized location. The files, which contain a virtualized server, are small in number (but large in size) and transferring them to a remote location is much easier than moving thousands of files.

Additionally, dealing with locked or open files is easier because a locked file is not a separate entity you can't rename or move; it is just a file nested inside the file with the virtualized server. Migration of open or locked files while users are accessing or writing to those files is not the key concern it would be with certain simple backup/recovery solutions. There is also not a concern over the security of those redundant files because if the lock is done properly, you won't be able to open and/or modify the file; thus, if the data inside the file is sensitive, you don't have to worry that when you store the file on a virtualized platform, the data inside it is readily exposed to unauthorized access.

The Benefits of an Effective Data Protection and Recovery Strategy

Through the flexibility of virtualization, you can encapsulate the OS, applications, and data into a virtualized system. You can then transmit this encapsulated application within your production environment or to an offsite location—just as you would transmit a data file—and make it available on a remote machine for employees to access. Even if there were no other benefits of virtualization in regard to data protection, this capability alone can cut downtime in the event of system failure from days to hours or less.

In considering the benefits of any disaster planning and recovery solution, there are two common metrics that are analyzed: Recovery Time Objective (RTO), which is a measurement of the time from the point of system failure and the time when that system is functional yet again, and Recovery Point Objective (RPO), which is a time definition connected with how much data loss is possible with the recovery. RPO relates to the timeframe of the last good backup that can be restored in the event of a total disaster. For example, if a disk array fails and data is only recoverable from your tape backups, which go back to the night before (roughly 10 hours for the sake of our example), your RPO is 10 hours. The interesting aspect of these two metrics is that there is the reality of what you have and the goal that you are reaching towards. Between the reality and the goal, there are a slew of technology options for high availability, clustering, hardware RAID, virtualization, and so forth. The tighter your RTO and RPO requirements, the more money you will have to spend, depending on the technology you choose.

Typically, companies are looking at their RTO and RPO results and seeing that traditional methods are not even close to being good enough for their desired objectives. This is what is driving the movement toward virtualization. Certainly the feature benefits mentioned earlier with the ability to perform testing and manage servers remotely and more easily are certainly a draw. But it is the ability to recovery quickly from a disaster with a minimal disruption of business operations at an affordable price where virtualization shines.