Seagate

Virtualization

Technology Paper

Disaster Recovery

Acronym: DR

The process, policies, and procedures that enable a business to recover data and systems after a disaster.

A Foundation for Disaster Recovery in the Cloud

Introduction

Virtualization has revolutionized the data center by changing the way it works. It is responsible for an entirely new ecosystem of products and serves as the foundation for cloud computing. In short, virtualization makes it possible to build and manage scalable, robust, resilient computing environments.

Combining a robust computing infrastructure with modern disaster recovery (DR) technologies and techniques makes it possible for service providers to offer scalable, efficient, and reliable DR-as-a-service (DRaaS) solutions in the cloud. An ever-increasing number of DR technologies are being built to complement virtual environments.

Beyond the technical merits of virtualization, there are economic benefits as well. This is particularly relevant in the context of cloud computing, and even more so when looking at the requirements for DRaaS.

In this paper, we explore why and how virtualization provides the perfect underpinnings for DR in the cloud. We will discuss specific virtualization features and different DR techniques, and investigate virtualization's economic benefits.

Virtualization

Virtualization

Hardware Abstraction

To create a virtual machine (VM), the software running on the host server (hypervisor) seamlessly abstracts the hardware. VMs run with virtual drivers, which enable the VM to communicate with the hypervisor. The hypervisor contains the drivers required for the specific physical server hardware on which the VM is running so the VM can run on different hardware free of any driver adjustments.

Because hardware adjustments aren't required for restoration after a VM image or snapshot has been taken of the full system—operating system, applications, and data—the practice of imaging VMs has grown rapidly. Gone are the days of copying only the data inside a system and then recovering the system by rebuilding it and trying to reattach the data. This highly error-prone process required an extensive amount of human intervention, resulting in long recovery times and, sometimes, unrecoverable systems. Full-system images eliminate the need to rebuild the operating system (OS) and reinstall applications, thus dramatically reducing errors and recovery times.

The hardware abstraction of a VM from the underlying server hardware means that not only can the same VM can run on different hardware without requiring no system changes required, as well as run it can also run on servers of different sizes, makes, and so on.

All this means that a DRaaS provider can build a recovery infrastructure without needing to consider the physical infrastructure.

hypervisor

Software deployed on a physical server/ host that enables the host to run virtual machines. Popular hypervisors include VMware® ESX and Microsoft® Hyper-V.

Snapshots

A snapshot is a point-in-time copy of a VM. One of the historical challenges in taking a server image was capturing the image from inside the still-running OS—the imaging technology had to work with locked files and keep up with changes to the system as the image was taken. Virtualization eliminates all that with snapshots. The snapshot is taken from outside the OS, removing the challenges faced by legacy imaging products.

P2V Migration Software

Notwithstanding the advantages of hardware abstraction and snapshots, organizations don't necessarily need to run virtualization to ensure effective DR. Physical-to-virtual (P2V) migration software has greatly improved the ability to image a physical system, making it possible to adjust an image of a physical system to run as a VM on a hypervisor. This means DRaaS providers can offer services beyond those that protect VMs only.

Network Isolation

Managing a multitenant, hosted environment requires making sure the various subscribers don't see each other's data. Typically, this type of isolation is managed through networking. On physical systems, managing individual network cards and ports on switches can be very time-consuming and error prone. On VMs, the networking layer is virtualized; most modern hypervisors provide easy mechanisms to ensure that different subscribers remain on isolated networks so there's no need to manage physical network infrastructure, even as VMs move from host to host or from storage to storage. This simplifies isolation and reduces the chances you'll accidentally expose subscriber data to another subscriber.

Other Considerations

Virtual environments provide numerous other benefits that can help ensure a more reliable and manageable environment. Perhaps the two most important benefits are:

- **Built-in resiliency:** If a host fails, its VMs can be automatically restarted on another host. This protects users and service providers alike from prolonged outages due to physical server failure.
- **Automated load balancing:** If one subscriber's VMs consume the bulk of a physical server's resources, the other VMs can move automatically to another physical host without causing any downtime. This ensures users can always access an environment that performs as expected regardless of other demands made on the host server.

Disaster Recovery Technologies

We've talked about why virtualization is an important technology for a service provider offering DRaaS. Now let's talk about the technologies that can be used to get data to the cloud. We will start by looking at replication and move on to talk about backup.

Replication

Replication technology typically captures changes in a production environment, and replicates them to a secondary environment where up-to-date standby systems can be run in the event of disaster.

Virtual Machine

Acronym: VM

A system deployed on a hypervisor with its own operating system, applications, and data. Multiple VMs can be deployed on one physical server.

Replication technology is commonly deployed via a software agent installed in the OS of the production system being protected. Because the replication software is in the OS, it doesn't need to know anything about the hardware. So it can run on any hardware using a supported OS, and it doesn't matter if the source and target environments are architected differently. This means service providers can build their environments on a virtualization platform regardless of whether the source environment is deployed physically or virtually.

Another more recent approach—specifically designed for virtual environments—replicates VM changes from a production environment to a secondary environment. This technology works outside the OS, so nothing needs to be deployed in the OS in order to protect the system. Another advantage: this purpose-built technology is a seamless extension for service providers wanting to leverage virtualization in their DRaaS offerings. But this technology tends to be limited to a single virtual platform, and multiple solutions are required to protect mixed environments—whether physical and virtual, or different virtual platforms.

A third approach is storage replication. Many storage area networks (SANs) can replicate data from one environment to a SAN in another environment. When integrated with the virtualization-based replication technology described above, SANs can be used to replicate VM system data.

Backups

Disk-Based Backups

The use of disk media is a key advance in backup technology. Relative to tape, disk provides significantly better recovery times and reliability. Disk backups are immediately available; tape backups must be physically retrieved from an offsite location and then mounted by an administrator via a time-consuming and error-prone process before recovery can begin. And in the past few years disk-based technology has become much more affordable, particularly with the advances in deduplication and compression.

It might seem like using disk for backup has little to do with virtualization. However, virtualization has greatly increased the use of shared disk storage, particularly SANs, which has driven down its cost. Nearly every modern data center has at least one shared disk storage device—and for that you can thank virtualization.

Long-Term Retention

Tape is still a very cost-effective media for long-term storage, but disk vendors and backup vendors are eroding that advantage by enabling disk to store more data at a lower cost. Many disk vendors use deduplication technology that makes it possible to store much more data on the same disk footprint. Disk-based backup vendors have invested in software-based deduplication, and many also run compression that can greatly increase the amount of data stored to disk.

Direct Boot

Direct boot means you can restart a VM directly from its backup. Traditional VM recoveries require software to transfer the VM backup to a data store (or some type of disk accessible to the hypervisor) before the VM can be powered on and recovered. These data-transfer times can vary widely depending on the size and number of the VMs.

Recovery Time Objective

Acronym: RTO

The amount of time a business can afford for a given system to be offline in the event of a disaster.

Direct boot technology bridges replication technology and backup technology. Traditionally, achieving a low recovery time objective (RTO) required using replication and creating a standby VM that could be powered on in the event of a disaster. This typically meant VMs needed to be both backed up and replicated, which required running two technologies.

Direct boot requires that a VM only needs to be backed up, not replicated. There is no need to transfer data, and thus no time lost during recovery—an advantage for the backup vendor and the customer.

A few important caveats, however, highlight the importance of evaluating desired RTOs before choosing a technology.

Replication typically still provides a lower RTO because the VM is running on native storage, whereas direct boot performance may degrade because the VM is running from backup storage. Backup storage is also often deduplicated and compressed, further slowing the recovery time slightly.

Replication and Backup Technologies Compared

As the chart below illustrates, there are pros and cons to using either replication or backup technologies.

Replication technology typically offers the lowest RTO and recovery point objective (RPO), making replication very attractive for DR. However, only backup technology provides the long-term retention required by certain regulations, and it can often provide granular recovery of individual items such as files or emails. Replication solutions also require a backup solution, which means a backup solution that can also be used for DR—thus obviating the need for a replication solution—is much less expensive.

For the service provider, virtualization makes it possible to build numerous solutions (including replication and backup) on the same infrastructure, and to offer different service-level agreements (SLAs) at different price points. DRaaS consumers can choose from multiple options that meet their specific needs.

Carefully analyze required RTOs and RPOs for the systems and applications in a particular environment when considering which technology to deploy or service to subscribe to.

Replication vs. Backup Technologies

	Replication	Backup
Definition	Copy system to secondary site to create a standby system	Copy data/systems to some type of backup media (tape, disk, cloud)
Primary purpose	Disaster recovery	Long-term retention, granular recovery
Cost	$$$ ($$$$$)	$$
Retention	Limited (hours/days)	Extensive (months/years)
RTO	Minutes	Hours
RPO	Minutes	Hours
Key Objection	Still require a backup product	RTO, RPO

Economics of Disaster Recovery in the Cloud

Economies of Scale

The term economies of scale means that the fixed cost of a good can be shared among all the goods produced—as the quantity increases, the cost per individual item decreases. Same for cloud-based DR solutions. If a company builds its own DR solution, it has to absorb the costs of building a data center, providing bandwidth, staffing the operations, and so on. A service provider, however, can share the fixed costs among all the subscribers, greatly reducing the costs of individual DR services.

Fractional Reserve

In the financial services industry, the term fractional reserve refers to the fraction of deposits a bank keeps on hand to satisfy the demand for withdrawals. DRaaS service providers can take advantage of virtualization to leverage this same practice.

Banks know only a fraction of their customers will withdraw money on a particular day; in fact, banks have developed algorithms to help meet the withdrawal needs of their customers. Similarly, service providers that are well distributed geographically know that not all their customers will experience a disaster on the same day. They don't need to purchase hardware to support the recovery of every server from every customer. However, they do need enough hardware to support all the recoveries required on any given day.

Recovery Point Objective

Acronym: RPO

The amount of data loss a business can afford for a given system in the event of a disaster.

Virtualization greatly aids in the creation of a pool of compute resources that can be easily consumed for one purpose, and then torn down and later consumed for another purpose.

The Business Case for DR in the Cloud

Let's compare the cost of in-house DR to a DRaaS subscription service.

In-house DR includes the following costs:

- **Hardware:** servers, storage, switches, racks, and other items required at the DR site
- **Software:** backup or replication software and licenses required to maintain DR site
- **Services:** hosting, bandwidth, setup, and so on—everything beyond hardware and software needed to build and maintain a DR site
- **Staffing:** head count—people required to build and maintain the DR solution and execute failover (tests and actual)

In-house DR costs include hardware, software, services, and staffing. The cost of maintaining the DR solution and environment, and of staffing when a disaster is declared or a test is executed, often gets overlooked. DR is an ongoing project that requires a (sometimes steady, sometimes surging) stream of investment of time, money, and effort.

DRaaS subscriptions include the following costs:

- **Subscription:** cost of the service
- **Staffing:** head count–people required to help test the DR solution

The cost for DRaaS tends to be very simple to calculate. Most service providers charge a subscription fee based on the number and size of systems being protected. The organization's staff will likely have some involvement in testing the DRaaS solution but the involvement tends to be minimal.

host

Synonymous with physical server; often refers to a physical server running a hypervisor.

DRaaS costs tend to be very simple to calculate: most subscription fees are based on the number and size of the protected systems. The organization's staff will likely be involved in testing the DRaaS solution, but the involvement is typically minimal.

Conclusion

Virtualization has made it possible to build a robust, high-performing, resilient, and cost-effective hosting environment. By leveraging virtualization, replication, and backup technologies, service providers can build a DRaaS environment that meets a range of customer needs. Virtualization also alters the economics of DRaaS, enabling service providers to profitably offer new services.

Take the Next Step

To learn more about EVault cloud services from Seagate, call us at 1.877.901.DATA (3282), email us at concierge@evault.com, or visit us at www.evault.com.

www.seagate.com

AMERICAS Seagate Technology LLC 10200 South De Anza Boulevard, Cupertino, California 95014, United States, 408-658-1000
ASIA/PACIFIC Seagate Singapore International Headquarters Pte. Ltd. 7000 Ang Mo Kio Avenue 5, Singapore 569877, 65-6485-3888
EUROPE, MIDDLE EAST AND AFRICA Seagate Technology SAS 16–18, rue du Dôme, 92100 Boulogne-Billancourt, France, 33 1-4186 10 00

2014.09.0012_wp_us (updated 09/30/2014)