IDC TECHNOLOGY SPOTLIGHT

Leveraging the Public Cloud for Faster Disaster Recovery at Lower Cost

May 2015

Adapted from *Disaster Recovery as a Service Builds Momentum as Businesses Reap the Economic Benefits of the Cloud Model* by Paul Hughes and Phil Goodwin, IDC #254455

Sponsored by Commvault

Disaster recovery (DR) does not get the attention it deserves from businesses because of the cost and complexity of deploying a full DR capability. Moreover, business units do not see a direct benefit from DR — until a disaster occurs. Traditionally, IT organizations duplicate their infrastructure in a second datacenter (i.e., doubling the infrastructure cost) or rent standby resources from a third party. Either solution is an expensive endeavor just to protect from an unlikely event. Yet, disasters do happen and not always to the other guy. Best practice organizations have robust DR capabilities that are regularly tested. This Technology Spotlight examines how cloud computing can be leveraged to develop DR capabilities that are both less expensive and easier to deploy than traditional methodologies. The paper also looks at the role of Commvault's cloud-based DR solution, Simpana, in the important market for DR services.

Introduction

The traditional model for establishing a disaster recovery plan for IT services requires at least two datacenters, each with identical infrastructure. In some cases, organizations use a specialized third-party datacenter that hosts multiple subscribers. One of the advantages of this approach is that a third party is responsible for maintaining the second datacenter. The disadvantages are that it's difficult to maintain hardware compatibility, the resources are costly and sit idly waiting for a disaster, and the systems may be shared with other organizations in the event of a regional disaster.

In other cases, organizations leverage two or more internal datacenters, again with duplicate hardware. The advantages are that both systems can be utilized (active-active) and are fully dedicated to the owner. On the downside, both sites must be significantly oversized to support a disaster, and facilitating failover is a remarkably complex process with respect to both technology and people. Either scenario involves significant costs to guard against a relatively rare occurrence. As a result, too few organizations give DR the attention that it deserves.

The emergence of cloud computing has introduced attractive new opportunities and capabilities for organizations that need a complete disaster recovery solution or are looking for ways to reduce costs without sacrificing service. In many cases, cloud infrastructure can be acquired "on demand" so that the organization will ultimately pay for only what it uses. With its cost advantages, cloud can make full-blown DR solutions available to small, medium-sized, and large businesses alike.

Despite the obvious benefits of cloud DR, it's not without challenges. Like all IT projects, it requires collaboration between business unit leaders plus careful planning. The biggest mistake organizations can make is to view DR as a purely technical exercise; DR is the classic triad of people, process, and technology. Recovery from a full disaster will require a coordinated effort across the entire enterprise, but it is an effort that can be driven by the IT group.

During the process of creating a robust cloud DR plan, IT managers should be prepared to deal with challenges such as the following:

- **Data staging** represents the critical path to DR. Without consistent, complete data, there will be no recovery. IT organizations need a simple, reliable method for transferring data from the production system to the cloud DR repository. Some cloud providers charge significant fees for data ingress and egress, so an efficient methodology for transferring data is required.
- **Service levels** must be established in coordination with business units to ensure that the proper DR capabilities are built to support business requirements. In general, the more stringent the SLA, the more costly it is to support. Senior managers need to understand the trade-offs between cost and performance so that the right balance can be struck.
- **End-to-end DR** must be considered, from compute to networks to storage to applications. Some orchestration mechanism is needed to coordinate the recovery of all components. Moreover, not all components will be identical. For example, there may be hypervisor differences (Hyper-V versus ESXi) and different storage systems. The orchestration layer must be able to bring together a wide variety of components.
- **Testing** must be conducted regularly to ensure the completeness of the solution, especially given that gradual change impacts infrastructure. Products that can simulate tests and do pretest checks are particularly helpful.

It is not possible to be a best practice organization without an adequate DR plan. Some key best practice activities are as follows:

- Know the organization's cost of downtime, at least for all tier 1 applications. This data can be used to cost justify any specific solution.
- Negotiate the recovery point objective and recovery time objective with business unit leaders, using the cost of downtime information to determine what is justifiable and what is not.
- Have a written DR plan that includes operational runbooks or automated processes (runbooks can now be automated) to ensure organizational survival in the event of a disaster.
- Test the DR system at least once a year (although quarterly is better, given that the proper testing tools are in place).

Cloud-based DR is a game-changing event for many DR plans, but it does not happen by itself. Organizations need the technological tools to integrate their on-premise systems with the cloud systems and to automate as much of the process as possible.

Definitions

To help organizations avoid the confusion that can result from different uses of terms, IDC offers the following definitions:

- **Recovery point objective (RPO):** The point in time at which data has been protected and can be restored with assured data integrity (Said differently, it is the maximum amount of data that will be lost, in terms of hours of operation, if a system failure occurs. For example, a one-hour RPO means one hour of lost data.)
- **Recovery time objective (RTO):** The total amount of time needed to restore application services from the moment those services are interrupted

- **Disk to disk to tape (D2D2T):** The process of copying data from a disk source to a disk target and subsequently on to a tape target for a total of three data copies
- **Disk to disk to cloud (D2D2C):** The process of copying data from a disk source to a disk target and subsequently on to a cloud target for a total of three data copies
- **Purpose-built backup appliance (PBBA):** An integrated unit of hardware and software used as a data backup target
- **Backup as a service (BaaS):** A cloud-based service where the cloud repository is simply a backup target, usually on-demand; does not include any additional compute or network resources for application recovery
- **Recovery as a service (RaaS):** BaaS functionality plus on-demand compute and network resources necessary to establish a functioning application environment
- **Disaster recovery as a service (DRaaS):** RaaS functionality plus consulting, runbooks, and personnel plans for extended application operations

Benefits

IDC estimates that 50% of organizations have inadequate DR plans and might not survive as a going concern after a significant disaster because of the inability to recover IT systems. In many cases, IT managers may have only a vague idea of how they would reestablish application services, lack up-to-date runbooks, and have no contingent personnel plan. While a DR budget from IT may be submitted annually to management, it always seems to be the first thing dropped from the overall budget as soon as constraints are applied; management often experiences sticker shock at the cost of DR for something perceived as a remote possibility. At the same time, business units may not be aware that an adequate plan does not exist and assume that operations would be restored promptly under any circumstances.

Calculating the cost of downtime is the best means of breaking this logjam. IDC research indicates that the average cost of downtime is about $100,000 per hour, although it can go as high as $1.6 million per hour for some organizations. This research also revealed that most organizations experience between 10 hours and 20 hours of unplanned downtime per year, even without a disaster event. This means that an average organization could spend between $1 million and $2 million on a contingency system and still achieve a one-year ROI. Given that recovery from a true disaster may take several days or even weeks without adequate planning, the cost-justifiable budget may be substantial. Cloud-based DR makes these economics even more attractive, with no huge up-front investment and variable pay-as-you-go operational expenses.

From a technology perspective, cloud offers a variety of data protection choices that build upon each other: BaaS is the basis for RaaS, and RaaS is the basis for DRaaS. Organizations can start with the basic BaaS and build toward full-blown DRaaS, assuming they choose a provider that offers that range of services. However, no matter how complete the provider's cloud solution might be, few providers offer products or services for the on-premise or private cloud portion of the customer's environment. Fortunately, solution companies are rapidly introducing the tools that IT staff will need to manage both the on-premise portion and the cloud portion of the DR solution.

Key Trends

Because of the compelling price-to-value proposition of cloud storage and DR, organizations are increasingly looking to cloud solutions for data resilience. IDC research estimates that the compound annual growth rates (CAGRs) for the BaaS and RaaS markets are 9.06% and 21.44%, respectively. Combined, the two markets will account for more than $1 billion by 2018 ($1.023 billion). Clearly, these markets have achieved critical mass and will attract the investment and competition that benefit consumers.

Cloud is also gaining acceptance well beyond the IT community. Business units commonly contract directly with software-as-a-service (SaaS) solutions (i.e., salesforce.com) and use cloud storage services for file sharing. Consequently, many end users are very comfortable with using cloud services and are supportive of employing cloud services for enterprise purposes. Some of this comfort, however, may lead to misperceptions by end users as to what is achievable in terms of service delivery. Given that a typical Web application is "always on" from their perspective, they will expect the same level of availability from the internal IT group. While this may be technically feasible, IT people know that always-on service delivery can be prohibitively expensive for single organizations. This is where the cost of downtime calculation and conversation with business unit leaders can help set appropriate expectations and gain buy-in from those leaders for an appropriate solution.

An unintended complication from this cloud storage scenario is that corporate data is becoming fragmented across service providers, from SaaS to BaaS to private cloud and so on. Regardless of location, this data is corporate data and needs to be managed and monitored by the IT group. In many ways, this can be a win-win for end users and IT alike. End users get best-of-breed solutions for their business needs, and IT is able to offload routine work to focus on more strategic objectives.

Considering Commvault

Commvault, a well-known data protection and information management vendor for more than a decade, is stepping up to extend its data protection capabilities to include the cloud. While it may be tempting for organizations to jump into the cloud data protection and DR space and build backward to the on-premise datacenter, it really makes more sense to extend the datacenter capabilities outward to include the cloud. This is a clear path and obvious opportunity for current Commvault customers, but it is also an opportunity for new customers to consolidate data protection across the enterprise regardless of where the data is located.

Simpana, Commvault's flagship product, provides the data movement, orchestration, and management capability to facilitate data protection and recovery both on-premise and in the cloud. Simpana provides a single point of control for replication, protection, and archiving of data to the cloud. Having a single "point of authority" for data protection not only is a best practice but also simplifies the administrator's efforts to recover data and results in faster RTO compared with a patchwork of solutions. Simpana coordinates backup images, snapshot images, replication, and virtual machine images to facilitate recovery of the entire application environment. This integration provides the management bridge to the cloud, bringing the various repositories into a cohesive entity.

As noted previously, staging data to the DR location is one of the most critical points in the DR process. In addition to managing data, Simpana can replicate data. Rather than have a separate replication tool as some solutions do, Simpana offers centralized management of backups, snapshots, and replication so that administrators don't need to search with multiple tools to find the right data copy. Commvault also offers a PBBA with cloud connectivity for organizations that want the convenience of an integrated software/hardware appliance as an easy gateway to copy data to the cloud.

The critical downfall of many DR plans is the lack of attention to the process of orchestrating the recovery of a production workload. Simply matching component for component between different physical locations is not enough. The process element of the recovery must be viewed holistically and begins from the time the data is created and protected for the first time.

Commvault's cloud DR solution uses Simpana as the recovery orchestration engine to give administrators a complete view. It is a vendor-neutral solution that supports major virtual machine (VM) environments such as VMware ESXi and Microsoft Hyper-V as well as major cloud services such as Amazon Web Services (AWS) and Microsoft Azure. This combination allows the entire application environment to be recovered in the cloud. Because both AWS and Azure offer on-demand compute infrastructure, organizations can establish test or recovery environments when they are needed and then deprovision them when the usage is complete. IT teams do not need to worry about maintaining device-level compatibility as they did with third-party DR subscription models in the past.

In addition, Simpana allows catalogs of prebuilt workflows to be created and utilized. From a DR implementation perspective, these workflows are essentially runbooks. Better than runbooks, however, is that the workflow is fully automated. For example, the workflow might include the creation of a specific virtual machine build and attachment to the related storage. In this way, an on-demand application environment can be established quickly and easily. Simpana is also VM aware so that newly provisioned VMs can be automatically detected and protected.

Though cost savings is the factor that receives all of the attention with respect to cloud DR (and it is real enough), agility is the other major factor that makes cloud DR attractive. The ability to move data between on-premise and cloud repositories, establish multiple DR sites if needed, and change providers as needed gives organizations the options to optimize recovery, convenience, and cost.

Because data used for DR is by definition moved offsite, best practices demand that the data be adequately protected from unauthorized access. Simpana's embedded encryption capabilities can be applied comprehensively or selectively by policy. These encryption policies apply enterprisewide, whether in the on-premise datacenter, public cloud, or hybrid cloud. Simpana also has roles-based access control integrated with Active Directory to simplify sign-on access for users while improving security.

Although cloud storage can be a very low-cost data repository at rest, some providers charge significant fees for data ingress and egress. To minimize these charges (as well as the amount of data that must be stored), Simpana embeds a flexible data deduplication mechanism. This deduplication can be applied to the source repository or the target, whether those repositories are in the main datacenter, in the cloud, or at a remote site. When data is deduplicated at the source datacenter prior to replication to the cloud, any ingress charges are minimized. Conversely, if the organization chooses to move the data from one cloud provider to another, deduplication in the cloud will minimize any egress charges. The current deduplication engine in Simpana is Commvault's fourth generation of the technology.

Even though DR capability is a business requirement, there's no reason that the repository needs to just sit idly waiting for something to go wrong. Commvault has been at the forefront of multiuse for backup images. With the capabilities of Simpana ContentStore, cloud, remote-site, or datacenter repositories can be used for user inquiries, eDiscovery, legal holds, and Big Data analytics. This multiuse of the data is a way for IT to deliver more value to business units without increasing costs.

Challenges

Simpana is very comprehensive in nature for backup and data protection but does not presently incorporate a DR "preflight check" or a test simulator. IT users will want to use either the traditional means of actual test failover or a third-party test simulator. The benefits of such a simulator are that it does not impact production operations and can often identify gaps or incompatibilities before any actual physical tests commence.

Conclusion

All organizations — but especially those without an adequate DR plan — will want to consider BaaS, RaaS, and DRaaS in the cloud. While disasters rarely occur, organizations with a cloud DR solution will most likely survive and thrive afterwards. Organizations that had been hesitant to deploy a full-blown DR solution because of the cost of traditional approaches can leverage cloud economics without huge investments while giving themselves many options should future circumstances change.

True disasters are thankfully rare, but outages are not. Most organizations suffer between 10 hours and 20 hours of unplanned downtime per year, and a cloud DR plan can be leveraged to recover from these more commonplace disruptions as well. When failover is fully automated, cloud recovery can be triggered at very low cost and deprovisioned when complete.

Commvault has leveraged its single, comprehensive Simpana platform to extend recovery capabilities into the cloud. With more than 80% of organizations expected to use some sort of cloud service by 2018, organizations obviously need a holistic data protection and disaster recovery solution. IT managers should learn from the lessons of the SAN/NAS era in which storage silos created management headaches; they should not allow cloud to become the new silo with entirely separate data protection and recovery schemes. DR is an organizational problem, not just an IT problem.

IT and business unit stakeholders should collaborate on implementation to ensure a corporate solution that meets business expectations and requirements. Site disasters are rare, but they do happen. Organizations must be prepared for the unthinkable.
