the Availability Digest

www.availabilitydigest.com

Recovery-as-a-Service

February 2013

Early adapters have proven the feasibility and cost advantages of moving applications to the cloud. Early on, cloud services such as Software-as-a-Service (SaaS), Infrastructure-as-a-Service (IaaS), and Platform-as-a-Service paved the way for many companies to take advantage of the cloud without having to get closely involved. SaaS offered CRM (Customer Resource Management) and ERP (Enterprise Resource Planning) services. IaaS provided web hosting services. PaaS let customers leverage the cloud providers' servers and storage systems for their own purposes.

Though most applications running in the cloud are still low-risk applications, more and more core applications are being moved to the cloud. This move is being aided by the provision of redundant cloud services such as Amazon's Availability Zones, which lets a backup copy of an application run in another fault-isolated Zone.

The new cloud killer-app may be Recovery-as-a-Service (RaaS), using the cloud to backup and to recover critical services that are running in a company's data center. Though RaaS has yet to become mainstream, cloud service providers, IT resellers, and startups are jumping on the bandwagon.

Traditional Disaster Recovery

Traditional disaster recovery is expensive and complex and often results in long recovery times and high data loss, both of which can be unacceptable to a company's operations. There are several considerations in a disaster recovery strategy for an application:

- RTO, or recovery time objective – This is the maximum amount of time for recovery that a company is willing to tolerate.
- RPO, or recovery point objective – This is the maximum amount of data that a company is willing to lose.
- Performance – The recovery solution must provide acceptable throughput and response time for every protected application.
- Consistency – Following an outage, the application must be returned to a consistent state to ensure the accuracy of the application.
- Geographic Separation – The production site and the backup site must be far enough apart so that no single disaster will impact both sites.

These needs have been met to varying degrees with various levels of backup:

- Cold backup site – Data is replicated periodically, such as every few hours or daily, typically to magnetic tape or to disk-based virtual tape, which is stored offsite for safety. Backup servers are not readily available. In the event of a site disaster, it can take hours or days to purchase replacement servers, to bring them out of storage, or to repurpose test or development systems for production use. Hours of data since the last backup will be lost. RTOs and RPOs are measured in days. This is the low-cost option for applications that do not require high availability.

- Warm backup site – Standby servers are kept in a standby state such that they can be put into service in a matter of hours. Periodic data replication moves production data to backup disks at the recovery site every few hours. RTOs and RPOs are typically measured in hours.

- Hot backup site – Servers at the backup site are kept mirrored with the production servers and are ready to be put immediately into operation. Data is replicated to the backup site every few seconds or minutes. RTOs and RPOs of minutes can be achieved.

These disaster recovery strategies have served the IT industry well for decades. However, they suffer from several challenges:

- They are costly. Not only does a remote disaster-recovery data center need to be built and manned, but duplicate server, storage, and networking equipment needs to be purchased and maintained (cold backups mitigate these costs to some extent).

- The cost of reducing RPOs and RTOs from days to hours to minutes becomes increasingly expensive.

- Backup data centers are expensive to scale incrementally. If more processing capacity is needed, new servers may have to be purchased.

- Backup data centers represent an increased burden on the company's IT staff to deploy, configure and administer.

- The integration, coordination, and scheduling of disparate systems at multiple sites is complex.

- The configuration of the production and backup systems must generally be identical. When a hardware or software change is made to the production system, it must also be made to the backup system. Configuration drift is a major cause of failover faults.

- Periodic testing of backup and recovery procedures is costly. It is often not done thoroughly, resulting in potential recovery failures following an outage.

Using Cloud Services for Backup and Recovery

Having dedicated infrastructure tied up waiting for a downtime event is a very inefficient use of resources. The cost and complexity of traditional disaster recovery sites has discouraged many small- to medium-sized businesses (SMBs) from implementing effective backup sites. Consequently, they are exposed to the dangers of a site outage. Government statistics show that 93% of companies that lose their data centers for more than ten days file for bankruptcy within a year.

Disaster Recovery for Small- to Medium-Sized Companies

Cloud-based disaster-recovery services have the potential to address these concerns.[1] With cloud-based RaaS, virtual images of physical or virtualized servers at the production site are maintained in the cloud, as is the current application data. Changes to application data or to server configurations are periodically replicated to the cloud to update these images so that the cloud images represent a reasonably up-to-date state of the production site. If the production site experiences a server, storage, or even a total site outage, the virtual images in the cloud can be invoked to continue processing in a time ranging from minutes to a few hours.

The primary concern for SMBs is the cost of disaster recovery. Cloud-based recovery addresses this concern directly. There are no upfront costs for a second data-center site or for duplicate servers, storage, and networking gear. Rather, backup images of virtual servers and of replicated data are stored in the cloud, but are otherwise passive, using just storage space and little processing. Cloud costs are on a pay-for-usage basis. Therefore, the cost during normal operation is relatively small. It is only in the event of a production outage that the virtual backup servers in the cloud spring into action and costs escalate. Hopefully, this is hardly ever.

A strong secondary concern is the load imposed on the existing IT staff to manage a disaster recovery site. With cloud backup, services are deployed and managed by the cloud provider. Processing facilities are easy to scale and provide protection to accommodate future growth. The virtualized backup servers and storage are easy to manage with automated cloud tools.

Cloud recovery services are positioned to move many SMBs to an effective disaster-recovery strategy.

Cloud-Based Recovery Options

There are several ways in which the cloud can be used to provide application backup services.

Both Production and Backup in the Cloud

The production applications can be run in the cloud, and backups provided in the cloud itself. This is the technique currently being used with, for instance, Amazon's Availability Zones. This requires a major commitment to cloud processing, but many large organizations are doing this for a range of non-critical to critical applications.

A major exposure with this method is the possibility of a total cloud failure. Though this is highly unlikely the probability of such an event is not zero. The Never Again stories in the *Availability Digest* are loaded with cloud failures lasting for hours or days, from Amazon Web Services and Microsoft's Azure Cloud to smaller cloud providers as well. If such a cloud failure occurs, the company's IT operations are down.

Backup to and Restore from the Cloud

The cloud can be used to replace magnetic tape or disk-based virtual tape. Used in this way, virtual server images and database changes are replicated periodically to the cloud. Should the company experience an outage, virtual machine images and up-to-date databases can be sent to a recovery site to continue production.

A limitation of this method is the recovery time of the database. Even modest-sized databases could take hours or more to transmit over the Internet or even a dedicated communication channel. This will significantly slow recovery.

[1] Much of the material for this article was taken from an excellent article by Nitin Mishra entitled Advantages of Disaster Recovery as a Service, published in *DataCenter Knowledge*.

Backup to and Restore in the Cloud

This is Recovery-as-a-Service, RaaS. With this technique, images of production physical and virtual machines are sent to the cloud, as is the application database. The virtual machine states and the application database in the cloud are maintained in synchronism with the production systems via replication.

If a server or the entire production site should fail, the corresponding virtual machines in the cloud can be quickly invoked, the user networks switched so as to reroute further traffic to the cloud, and application services can continue.

Prioritizing Applications

The amount of failure protection needed by a company varies with the application. Some applications can tolerate days of downtime, and no backup facilities may be necessary. Others can be down for a day, for a few hours, or perhaps for only a few minutes. This is often related to the cost of downtime, which may range from negligible to seven figures per hour.

The requirements for each application must be specified, and an appropriate backup strategy for each determined. This determination will be a balance between the cost of downtime and the cost of recovery.

Note that it is not really the application availability that is important. It is the availability of the services provided by the application that counts. If there is an alternate way to continue providing critical services should an application fail, this may affect the recovery strategy for that application.

The Advantages of RaaS

The advantages of RaaS are many. We have already described one of the most important advantages – the significantly reduced costs of creating and maintaining a backup facility. Other advantages include:

- Backup resources can be provisioned in minutes. RTOs measured in minutes or an hour or so are quite achievable. Faster recovery time (seconds to minutes) can be achieved by having the backup virtual servers up and running, ready to take over. However, this obviates to some extent the cost advantages of RaaS as these resources will have to be paid for even when the production system is operating.

- The backup services can be activated from a user portal – perhaps a wirelessly connected laptop. There is no need to call the cloud provider. The backup process can begin as soon as an outage is detected.

- The backup and recovery process can be automated, lowering recovery times and recovery failures.

- Virtualization eliminates hardware and software dependencies because the cloud's virtual environment is kept synchronized via replication. Configuration drift is not a concern.

- The backup configuration can be easily scaled by user command. If more processing power is required, it can be configured remotely by the user. If less processing is required, virtual machines can be shut down to save costs. Resources are elastic.

- The cloud provider manages and maintains the disaster recovery services. This reduces the customer's ongoing IT costs.

- The cloud disaster recovery system is easier to test. Tests can be run quarterly or monthly rather than annually,

Challenges with RaaS

Recovery-as-a-Service does not come without concerns.

- The same effort for preparation, planning, and testing that would normally accompany a warm-site build-out is required for a RaaS implementation.

- Staff must be trained in failover and fallback procedures.

- The location of the cloud facilities should provide the geographical separation from the production site required to prevent a common disaster from affecting both.

- The company has less control over transaction throughput and response time. It must be ensured that enough cloud resources can be provisioned to provide the capacity needed by the production applications.

- In bringing up the backup resources, there will be interdependencies between applications. Applications must be brought up in the proper order.

- There is a dependency on network availability and capacity. Dual networks to the cloud provider may be required to ensure connectivity when needed. If the Internet is to be used, will it support the required response times? Unanticipated extra bandwidth may take time to provision and will result in extra costs.

- Network procedures must be put into place to redirect users, employees, and IT staff to the cloud-based services.

- The cloud provider's services must comply with regulatory security requirements such as those for financial and health services. Encryption at rest and in-flight may be required.

- There must be a procedure for failing back to the production site when it is returned to service. This may require a phased recovery of the production services.

- Ancillary services may have to be addressed. For instance, an application may need to print documents for the users.

- Cross-DR providers may create a particular problem Most clouds are designed to handle Windows, Linux, and Unix virtual machines. This is a function of the virtualization facility that they are using. However, many mission-critical applications run on mainframes whose backups are likely to be located in other backup data centers. Provision must be made to integrate the virtualized cloud backup site with these other DR sites.

- The disaster recovery site and failover procedures must be tested periodically.

- If a major disaster should affect many of the cloud provider's RaaS customers, the provider may not have enough capacity for all customers.

Evaluating a RaaS Solution

When evaluating a RaaS solution, several aspects must be considered. Replicas of all protected systems and data must be replicated frequently enough to satisfy the applications' RPOs. Systems should be

recoverable in a time sufficiently short to meet the applications' RTOs. The recovery of any element (file, server, disk, storage system) should be controllable by a user-driven self-service portal. Communication facilities between the production site and the cloud should be optimized to ensure full data mobility at minimum bandwidth. The cloud must comply with the company's security policies.

The SLA with the cloud provider should address the following points:

- The lead time to allocate minimum required resources following a disaster.
- The lead time to scale up to required capacity.
- The duration for which resources will be retained on a dedicated basis for the company.
- Additional fees for occupancy beyond the defined period.
- The costs for additional facilities such as conference rooms, video conferencing.
- Provisions for failover and fallback testing.
- Cloud-based validation of replicated data (for instance, periodically comparing block checksums).
- The capability to provide additional capacity if needed.
- A guarantee that resources will be available in the event of multiple data center failures of other customers of the cloud provider.
- The required data security policies.

As a final step before signing on with a cloud provider, a company should run drills verifying that the company's recovery requirements can be met.

Summary

Implementing a disaster-recovery solution is never simple. Cloud-based recovery can make the task much easier, more reliable, and less costly for many companies.

Large companies already have disaster-recovery infrastructures in place and may not be so ready to move recovery to the cloud to protect their valuable and critical services. Small companies are less likely to have a formal DR strategy.

However, cloud recovery will become attractive, especially to medium-sized companies. Gartner predicts[2] that 30% of mid-sized companies – those with annual revenues from $150 million to $1 billion – will have adopted RaaS by 2014. That is up from 7% in 2011.

[2] Gartner Says 30 Percent of Midsize Companies Will Use Recovery-as-a-Service by 2014, *Gartner*; November 7, 2011.