Leveraging the Cloud for Data Protection and Disaster Recovery

Bennett Klein

DATA MANAGEMENT CUSTOMER SOLUTIONS

MARCH 2012

Table of Contents

Executive Summary

Summary

Executive summary

The “cloud” can mean many things; it can refer to applications and platforms that are delivered online, or infrastructure services such as data storage and server processing using remote data centers. Cloud services can be provided and delivered remotely as “public clouds” by vendors such as Amazon or Microsoft, or “private clouds” where the resources are created, deployed, managed and controlled in-house. Technologies that exploit the cloud are becoming increasingly important, providing tools and platforms for rapid web-based application development, distributed network infrastructures and secure offsite data storage.

According to a North America cloud survey conducted by Coleman Parkes in February 2012, over a third of companies (36%) in the U.S. report using a public cloud for data protection or applications, and in Canada, 28% do the same. Many businesses lack the internal resources (for example, remote sites, equipment and staff) to meet demanding service-level agreements (SLAs) for business continuity (BC) and disaster recovery (DR). Therefore, online cloud-based Infrastructure-as-a-Service providers can be used as an often lower-cost option than recruiting and retaining such skilled workers, as well as incurring large capital expenditures, and can provide services to meet SLAs. Often, organizations do not have their own remote facilities for DR, and “renting” remote resources from cloud vendors is a cost-effective alternative option. Cloud services can help where there are limited capital expense budgets, by transferring costs to annual operating expenditure that can be managed and only paid for as required or consumed, in a pay-per-use model.

This paper provides recommendations on how to approach the use of cloud resources for system, application and data protection. It looks at how you may use the cloud as a backup, replication and archiving location, and for providing a resilient offsite platform for enabling high availability of critical systems.

Section 1: Cloud technologies

There are many “cloud offerings” available today, from simple applications such as Gmail or Salesforce.com, to computing infrastructure resources provided across the Internet, such as cloud-enabled storage and cloud-hosted virtual servers. Cloud-based applications are typically referred to as Software-as-a-Service or Application-as-a-Service. Computing, storage and accessibility services provided across the Internet are referred to as Infrastructure-as-a-Service; where storage is the main service, this may be referred to as Storage-as-a-Service. Some vendors provide rich computing platforms, which are typically called Platform-as-a-Service. Examples include Salesforce.com and Amazon Web Services™ (AWS).

This paper focuses on Infrastructure-as-a-Service and Storage-as-a-Service and will focus on the use of public cloud infrastructure (such as Amazon Web Services and Microsoft Windows® Azure™) for offsite data protection and system availability.

One of the benefits of using a public cloud infrastructure service is that IT organizations can subscribe to these services in an “on-demand” manner and can scale up or scale down as necessary. Both Amazon Web Services and Microsoft Windows Azure leverage virtualization technologies to offer their customers this level of scalability and agility.

Amazon Web Services (AWS) offers a suite of cloud services. For data and systems protection, the following AWS services are most significant:

- **Amazon Elastic Compute Cloud (Amazon EC2)** provides online hosted virtual machines (VMs) that are used either as your primary or production servers or are used as failover servers for business continuity and disaster recovery purposes. More information may be found at http://aws.amazon.com/ec2/.
- **Amazon Simple Storage Service (Amazon S3)** is an online storage service that can be accessed via the web. Similar to Amazon EC2, Amazon S3 can be used as your primary storage infrastructure but is typically used for offsite data protection, archiving and disaster recovery purposes. More information may be found at http://aws.amazon.com/s3/.
- **Amazon Virtual Private Cloud (Amazon VPC)** provides secure virtual networking services, and is required for accessing Amazon EC2 machines. More information may be found at http://aws.amazon.com/vpc/.

Microsoft Windows Azure includes online storage and application fabric services that can be used as your production environment or as a remote site for business continuity and disaster recovery. The following Windows Azure services are most significant:

- **Windows Azure Compute** provides an application processing environment service where IT organizations can run their applications in a production scenario or as a failover environment in case their onsite resources become unavailable. More information may be found at http://www.windowsazure.com/en-us/home/features/compute/.
- **Windows Azure Storage** is an online storage service that can be used as primary or secondary storage, typically used as part of a disaster recovery or archiving strategy. Storage is offered as Binary Large Object (BLOB) and Table formats to meet different needs. More information may be found at: http://www.windowsazure.com/en-us/home/features/storage/.

Section 2: Key considerations for using the cloud

For many organizations, there are important issues that must be addressed as a priority while considering the cloud as a protection and recovery option:

- **Service-level agreements (SLAs)**. Most organizations will require SLAs to be in place to govern the availability of the protected systems and data, and guard against unplanned outages, service disruptions and data loss in the cloud provider's data center.

- **Legal compliance**. Whenever data is stored on remote servers, there are legal standards and compliance issues to resolve. For example, the geographic location of cloud data centers can be a critical issue because regulations in one country may be stringent and enforced more strictly than those elsewhere.

- **Manageability**. Many businesses perceive that a move to the cloud will involve a loss of control of their systems and data; the choice of cloud vendor is therefore very important.

- **Security**. Any systems and data copied to the cloud must be provably secured, but also remain accessible at all times. Security must be considered at all levels, including protecting data in the network as it is transmitted to the cloud data center, as well as within the data center after the data is stored.

- **Bandwidth**. For many organizations, available wide-area bandwidth and its cost considerations are significant barriers to the adoption of cloud-based system and data protection and recovery services. Careful planning is required to ensure that you have adequate bandwidth for both offsite protection and recovery back to the data center. The time that it takes to transmit data to the cloud and back for recovery needs to be tested and understood before turning to the cloud for IT delivery.

- **Long-term retention policies and archiving**. The cloud can help meet requirements for long-term data storage, as long as the remote data remains accessible and archiving policies can be simply and easily enforced.

- **Recovery testing and validation**. Any data and systems protection strategy is judged on its ability to deliver rapid and reliable recovery. Therefore, it is essential that recovery from cloud-based resources can be tested and validated routinely before the need for such recovery occurs.

The cloud can enable a good archiving and disaster recovery resource. It can even be a failover location for high availability of your critical applications and data for business continuity, as long as there is careful planning and the right tools are used. The cloud may provide an IT organization with more agility and flexibility in deploying its BC/DR strategy and resources, while helping to reduce costs by reducing the need to purchase, deploy, manage and maintain the associated hardware and software. It may also be the perfect solution for companies that do not have their own remote DR site or data center to use. Using the cloud may help IT organizations better balance their capital expenditure (CAPEX) and operating expenditure (OPEX), because the cloud is typically licensed as a monthly recurring charge on a "pay-per-use" model.

Section 3: Using the cloud for backup and recovery

Backup and recovery refer to the protection of critical data and applications across your IT environment, such as user files and folders on file servers, Microsoft Exchange mailboxes and Microsoft SQL Server® databases.

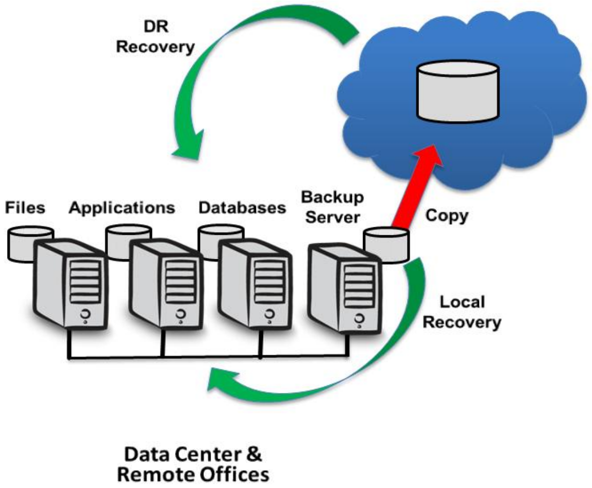
Today, many commercial backup solutions offer the use of cloud storage resources for offsite data protection and archiving. There are hardware, software and SaaS solutions offering file-based or image-based backup technologies. Most of these solutions are designed to protect physical and virtual servers while others can be used to protect workstations and laptops too.

The cloud can provide a good offsite storage resource for backup and archiving, but there are two primary challenges to consider:

- Backup windows. Depending on the volume of data or the size of a database, it will take far longer to perform backup directly to a remote cloud resource than to a local resource, which means that backup across the wide area network (WAN) can take longer than your current backup window allows.
- Recovery times. Similarly, it will take longer to recover data or a database directly from the cloud than from local disk. Some vendors even offer physical media transport with next day delivery to compensate for this challenge. The question to ask yourself is, can you wait up to 24 hours to recover? Can you meet your recovery time objective (RTO) and SLA? And how will you recover a complete system in case of an unplanned system outage or hardware failure?

A better approach is a **hybrid** model. This is where backup is performed locally for faster backup performance, and then a copy of the backup or critical files can be sent to the cloud afterwards, when performance is not as critical. Having a hybrid solution means faster recovery time, too, because you have the option to recover the data from a local source. The cloud is better left for offsite data storage for disaster recovery and file archiving purposes.

Fig. 1: Hybrid-cloud data protection example



In this hybrid-cloud backup example, backup is first performed using onsite resources to address demanding backup window constraints. Once local backup is completed, an automated copy to the cloud is performed for disaster recovery and archiving. IT organizations can apply file versioning and retention policies to help reduce onsite and cloud storage costs and address compliance requirements. Data can also be compressed, deduplicated and encrypted to achieve lower storage costs and data security goals. If data is lost or damaged by accident, malicious event, or even system outage, recovery is performed using the onsite backup storage for fastest recovery time. Recovery from the cloud would only be performed when the data center experiences a prolonged outage and needs to recover from the remote storage—hence true disaster recovery. Using a hybrid-cloud data protection solution means you can use a bare metal recovery (BMR) technology to help slash system recovery time over traditional methods.

While it might be less costly to perform backup directly to the cloud since you avoid onsite storage costs, a hybrid-cloud data protection solution will generally deliver faster backup, restore and system recovery along with offsite data protection for disaster recovery. Having a copy of your backups onsite by using a hybrid approach also decreases the risk and impact to your business in the unlikely event that your cloud service provider experiences their own server crash or data center outage.

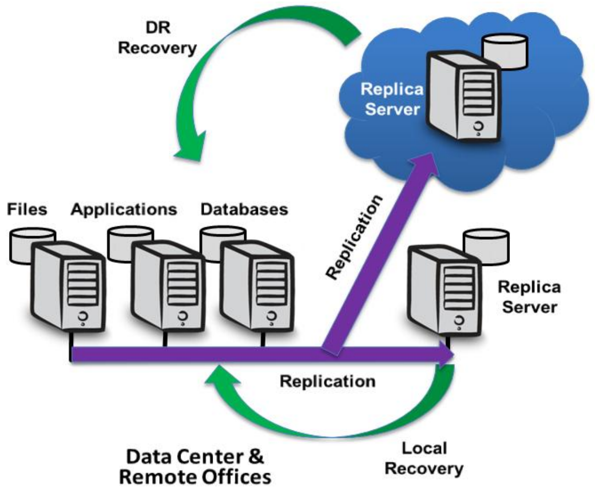
It’s critical that you research and understand the service level agreement being offered by your cloud provider and their disaster recovery plan as both will affect your own SLAs and DR plans. Ask if they have system high availability solutions and if they have multiple datacenters and perform replication of your data between them. You may want to consider using a cloud service provider whose data center is geographically located far from your own data center to avoid the possibility of local disasters affecting both you and them.

Section 4: Using the cloud for replication and CDP

In its basic form, replication is used to copy backups offsite for disaster recovery, after onsite backup has completed. But many organizations leverage replication as a continuous data protection technology to complement periodic backups, especially for more critical data. Such replication is performed in a real-time continuous manner, capturing each and every change made to systems, files, data and databases. This provides better protection in case of unplanned data loss and damage—especially when a storage device fails. Many organizations perform replication to a remote or offsite facility like the cloud to address both demanding recovery point objectives (RPOs) and disaster recovery strategies, but what can you do if you do not have an available remote site? The cloud can be a perfect vehicle for offsite replication, especially if you do not have your own remote facilities and staff.

More comprehensive replication solutions include a “data rewind” capability that offers a continuous data protection (CDP) solution where files, data and databases can be rewound back to a known good-point-in-time before the data loss or damage. This CDP technology is typically only used after unplanned events that occur between periodic backups as all data after the rewind point is typically lost.

Fig. 2: Hybrid replication example



Typically, most host-based replication software solutions are asynchronous to help overcome high-latency network challenges typically used to copy data to a remote cloud facility. In this example, production storage and replica storage are synchronized, and only byte-level changes are replicated across the network, whether onsite or in the cloud.

Just like backup, replication can be performed in a hybrid approach where data is replicated to low-cost secondary storage located onsite **and simultaneously** replicated offsite to the cloud. This method is referred to as one-to-many replication and helps you address both recovery point objectives (RPO) and disaster recovery strategies. Of course IT organizations can replicate directly to the cloud if preferred and thereby avoid additional onsite server and storage resources and costs.

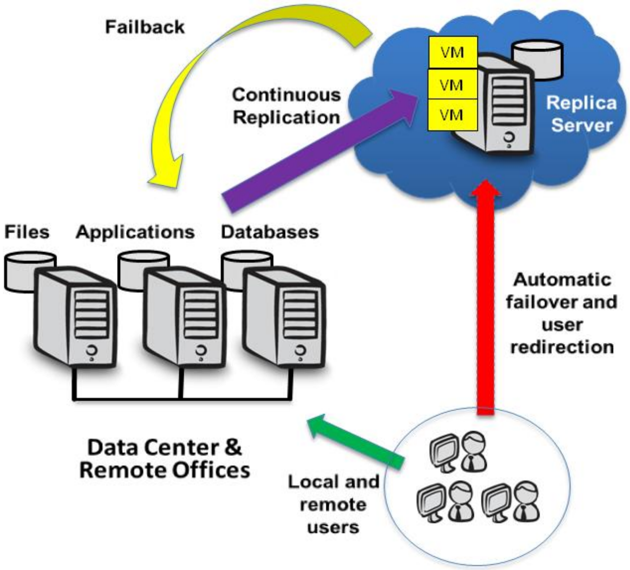
Section 5: Using the cloud for high availability

Whether you lose a critical server or your entire data center, restoring a full system, application and data (or database) after a system or storage failure can take many hours, affecting all parts of the business, and with potentially disastrous impact on sales and services. Employee productivity and morale are also negatively affected, as is reputation and even compliance.

To avoid business disruption while recovering your production servers or entire data center, IT organizations can replicate their full system image-based backups to the cloud, and after an outage, “stand them up” in cloud virtual machines and then redirect end users so they can get back to work. After production systems are repaired or replaced, an image backup of the cloud storage would be taken and then used to restore the local systems. End-users would then be redirected back.

For mission-critical systems, applications and data, a host-based, high-availability (HA) software solution may be necessary. This solution protects physical and virtual servers and storage, and may be deployed onsite, offsite and in the cloud. Although many organizations deploy high availability as an onsite solution for business continuity, it is also common to deploy this solution offsite or in the cloud to support both business continuity and disaster recovery needs by using the same technology. For organizations that do not have their own DR site or remote facilities to use, using a public cloud may provide an ideal solution.

Fig. 3: System and data high availability example



Host-based HA software solutions are typically asynchronous to help overcome high-latency network challenges in replicating systems and data to a remote site or cloud facility. Initial full system synchronization must take place between the onsite production server(s) and storage, and the cloud-based virtual machines (VMs) and storage (Replica). The HA system continually captures all byte-level changes and sends them to the Replica in the cloud keeping data and databases current. The HA solution performs real time system and application monitoring and once it detects a system or application outage, it automatically fails over to the cloud-based system (Replica) and automatically redirects end-users. Alternatively, high availability solutions may be set for "manual failover" where once an outage is identified by the IT organization; they can choose whether to failover and, if so, use push-button failover if failover to the cloud is desired. Manual failover can also be used in the case of an impending natural disaster or planned outages (e.g., for maintenance).

After the production servers are repaired or replaced, a "fail-back" is performed that will re-synchronize the Replica and onsite production server and redirect end-users back to the normal production environment. In many cases, use of a virtual private network or virtual private cloud solution will be required to access remote systems and applications.

Summary

A public cloud can be an important component of your BC/DR strategy, especially if you don't have your own remote site. You can use it for offsite backup and long-term data storage, as a secure offsite host for CDP and even for failover servers to achieve a high-availability environment.

About the Author

Bennett Klein is Senior Director of Product Marketing for CA Technologies where he is responsible for setting strategies for building worldwide market awareness and demand for the CA ARCserve Family of Products.

This document is for your informational purposes only. CA assumes no responsibility for the accuracy or completeness of the information. To the extent permitted by applicable law, CA provides this document “as is” without warranty of any kind, including, without limitation, any implied warranties of merchantability, fitness for a particular purpose, or noninfringement. In no event will CA be liable for any loss or damage, direct or indirect, from the use of this document, including, without limitation, lost profits, business interruption, goodwill, or lost data, even if CA is expressly advised in advance of the possibility of such damages