DISASTER RECOVERY IN CLOUDS

Sindhu S. Pandya*

ABSTRACT

Cloud computing has been a new buzzword of today's most enticing technology areas due to its cost-efficiency and flexibility. It offers mainframe or better infrastructure through a small set of services delivered globally over the Internet. As per one of the estimates from Gartner, by year 2012, 20% of enterprise market e-mail seats will be delivered via Cloud.

Many businesses rely on Disaster Recovery (DR) services to prevent either manmade or natural disasters from causing expensive service disruptions. Unfortunately, current DR services come either at very high cost, or with only weak guarantees about the amount of data lost or time required to restart operation after a failure. Cloud computing platforms are well suited for offering DR as a service due to their pay-as-you-go pricing model that can lower costs, and their use of automated virtual platforms that can minimize the recovery time after a failure.

In this paper, I focus on cloud data storage security and describe what challenges remain in order to minimize cost, data loss, and recovery time in cloud based DR services. I had performed a pricing analysis to estimate the cost of running a public cloud based DR Service and show cost reductions compared to using privately owned resources.

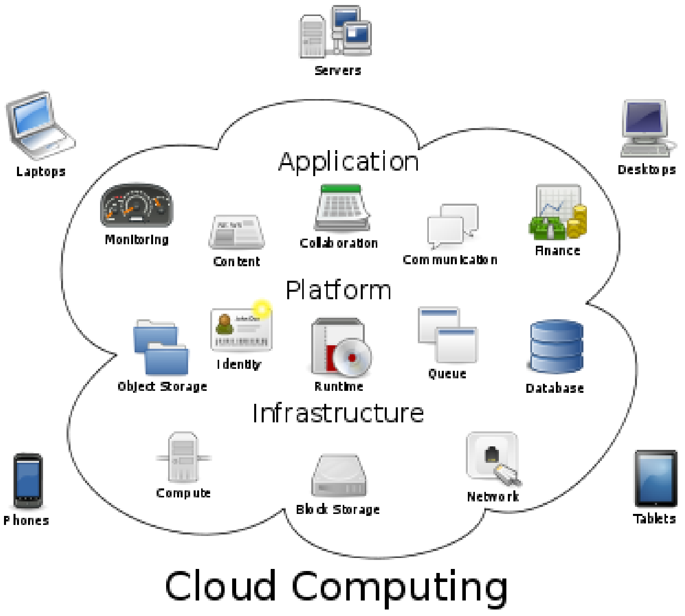
Keywords: *Cloud computing, security, privacy, recovery.*

*I/C Principal, Laxmi Institute of Computer Applications(BCA), Gujarat

1. INTRODUCTION

Cloud computing is a computing paradigm, where a large pool of systems are connected in private or public networks, to provide dynamically scalable infrastructure for application, data and file storage. Our society's growing reliance on computer systems means that even short period of downtime can result in significant financial loss, or in some cases even put human lives at risk. Considering the case happened on February, 5th 2011, in Tel Aviv, Israel, in the early morning hours, a massive fire broke out at the local IKEA store. Dozens of fire-fighters rushed to the spot to put out the flames. Some fire-fighters entered the burning building to check if there were any people trapped inside, members of IT team that worked there accompanied them hoping to save some of the computing gear and precious data before it was lost in the fire, thinking that the data was more precious than their lives. Keeping in view these factors many business and government services utilize Disaster Recovery (DR) systems to minimize the downtime incurred by system failures.

The main challenge in providing DR services is to allow applications to rapidly come back online after a failure occurs, but at high cost. The "pay-as-you-go" model of cloud platforms can lower the cost of DR. Our goal is to show how cloud computing can provide low cost DR services by minimizing data loss and recovery time.



2. CLOUD COMPUTING BENEFITS

a) **Reduced Cost:** The billing model is pay as per usage; the infrastructure is not purchased thus lowering maintenance. Initial expense and recurring expenses is also low.

b) **Increased Storage:** With the massive Infrastructure that is offered by Cloud providers today, storage & maintenance of large volumes of data is a reality.

c) **Flexibility:** With enterprises having to adapt, even more rapidly, to changing business conditions, speed to deliver is critical. Cloud computing stresses on getting applications to market very quickly, by using the most appropriate building blocks necessary for deployment.

3. CLOUD COMPUTING CHALLENGES

Despite of the growing influence of cloud computing, there are many drawbacks. Some common challenges are:

a) **Data Protection:** Enterprises hesitate to buy an assurance of data security. They fear losing data and data confidentiality of consumers. The actual storage location is not disclosed, adding onto the security concerns of enterprises. In the existing models, enterprises own firewalls across data centers protect this sensitive information. In the cloud model, Service providers are responsible for maintaining data security and enterprises would have to rely on them.

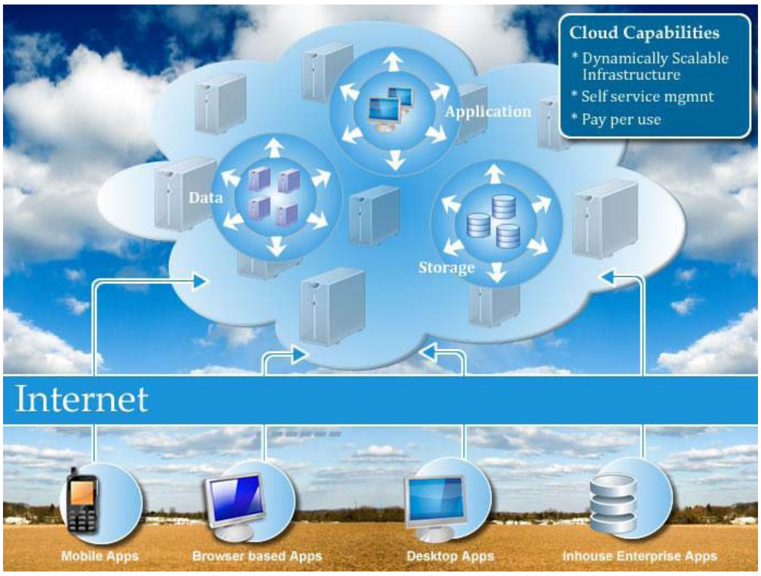
b) **Data Recovery and Availability:** All business applications have service level agreements that are followed. Operational teams play a key role in management of service level agreements.

- Appropriate clustering and Fail over
- Data Replication
- System monitoring
- Maintenance
- Disaster recovery
- Capacity and performance management

c) **Management Capabilities:** Despite there being multiple cloud providers, the management of platform and infrastructure is still in its infancy.

d) **Regulatory and Compliance Restrictions:** In some of the European countries, Government regulations do not allow customer's personal information and other sensitive information to be physically located outside the state or country. In order to meet such requirements, cloud providers need to setup a data center or a storage site exclusively within the country to comply with regulations.

To meet all these challenges of the present cloud computing we come to Cloud DR.



4. ENTER CLOUD DR

DR Service works by replicating application state between data centers; if the primary data centre becomes unavailable, then the backup site can take over and will activate a new copy of the application.

4.1 Requirements for DR

On the basis of business decisions DR requirements are as follows:

Recovery Point Objective (RPO): The RPO of a DR system represents point in time of the most recent backup prior to any failure. The RPO is business decision for some applications is that no data should be lost (RPO=0), requiring continuous replication to be used, while in others, the acceptable data loss could range from a few seconds to hours or even days.

Recovery Time Objective (RTO): The RTO specifies the time period for an application to come back online after a failure occurs. This includes the time to detect the failure, prepare servers in the backup site, initialize the failed application, and perform the network reconfiguration required so that the application can be reused. For a good business continuity we require very low RTO.

Consistency: The DR service must ensure that after a failure occurs the application can be restored. It is required that DR to be application specific to ensure that all relevant state is properly replicated to the backup site.

Geographic Separation: It is important that local and backup sites should be geographically separated so that both will not be affected at the same time by a single disaster. This long distance geographic separation incurs to higher WAN bandwidth costs and will incur greater network latency, which directly increases response time in case of synchronous replication.

Asynchronous replication techniques can be used for long distances, but can lead to greater data loss during disasters.

4.2 Working of DR

DR is a long distance state replication with the ability to start up applications at the backup site after a failure is detected. Replication can be done at one of the following layers:

(i) Within an application

(ii) Within a file system

(iii) Full system context

Replication within an application can be done only by transferring the crucial state. Replication at the file system duplicates all or a portion of the file system to the remote site. The level of data protection and speed of recovery depends on the type of backup mechanism used and the nature of resources available at the backup site.

4.3 DR services:

1. **Hot Backup Site:** A hot backup site provides a set of fully powered stand-by servers that are always available to run the application once a disaster occurs, providing minimal RTO and RPO. They use synchronous replication to prevent any data loss due to a disaster. They are very expensive.
2. **Warm Backup Site:** A warm backup site provides standby servers to run the application after a disaster occurs, but are only kept in a "warm" state that is it may take a short time to bring them online. This slows recovery, but also reduces cost. It may use either synchronous or asynchronous replication schemes depending on the necessary RPO.
3. **Cold Backup Site:** In a cold backup site, replication of data is done on a periodic basis, leading to an RPO of hours or days. In this servers which are needed to run the application after failure are not readily available, they have to be brought out of storage or repurposed, resulting in a high RTO. They are very cheap.

Cloud computing means that it provides the greatest cost benefit when peak resource demands are much high. Cloud platforms can provide the greatest benefit to DR services that require warm stand-by replicas. The cloud can be used to cheaply maintain the state of an application using low cost resources. Only after a disaster occurs cloud based DR service be paid for the powerful and expensive resources required to run the full application.

4.4 Failover and Failback

A DR solution must be able to detect when a disaster has occurred which is a challenging problem since temporary failures can trigger false alarms, then perform a failover procedure to activate the backup site, also run failback steps necessary to revert control back to the primary data centre. Most DR techniques rely on manual detection.

5. THE BUILDING BLOCKS OF CLOUD BASED DR

The first step is to perform a business impact analysis, defining which business processes are most critical and impacted by application downtime, and then applications are prioritized and organized.

5.1 Modernized data movement technologies

Selecting the right data movement technology is essential for an organized, efficient and successful recovery. For example, traditional tape-based backup delivers recovery anywhere from two days to a whole week. For critical applications like e-mail and online order systems, that level of downtime could shutter many businesses for good. Online technologies, on the other hand, can help get applications back online in a matter of hours.

Three categories of technology modernize data movement for application recovery are: server replication, storage replication, and vaulting.

Server replication is used for critical applications with a smaller numbers of servers. Storage replication is also a right solution for critical applications, and is best deployed with larger numbers of servers. Vaulting, also known as online backup and recovery, eliminates the cost and delay of tape-based backup.

5.2 Selecting a service provider

Each of these solutions can be deployed on the cloud. The fundamental benefits of moving recovery to the cloud are lower cost and scalability. But for critical applications, security and reliability become the fundamental benefits.

6. CLOUD – BASED RECOVERY MODEL

Every IT applications have specific recovery requirements. The following model illustrates how a fully managed solution applies a range of technologies to help ensure recovery requirements ensuring security, reliability, and cost-effectiveness.

6.1 Selecting the technology

With cloud-based recovery, once the business impact analysis is completed, then applications and data can be protected on a shared, secure cloud using the most appropriate data movement technology.

For the most critical applications, server replication are done for their ability to deliver recovery in less than four hours at RPO = 0. Server replication software encapsulates the OS, applications configurations and applications data, and replicates them as soon as changes are written to disk. Those applications that are less sensitive are encrypted and stored in a secure offsite location for online backup and recovery.

6.2 Handing over recovery responsibility

With a fully managed solution, the service provider is now responsible for the management and maintenance of the full recovery lifecycle.

During normal operation, the service provider executes 24/7/365 monitoring and handles monthly capacity tuning. In case of any disaster, the service provider will recover applications. The service provider coordinates the recovery in concert with additional applications on physical systems and legacy mainframe equipment, and reconnects business users to the recovered applications and data. When the incident has passed, the production environment is then returned to normal operations.

Managing recovery effectively is complex. A fully managed recovery solution removes the burden of that complexity from an IT organization. And when the recovery solution leverages modernized data movement technologies and a range of secure cloud and virtual infrastructures, the advantages are significant: higher availability, better performance, and lower cost.

7. SUMMARY: CLOUD STORAGE TODAY, DR TOMORROW

When it comes to Cloud-Based DR, organizations seem to be confusing on the storage aspect. In a global survey conducted last year, research firm Aberdeen found that organizations that utilized public Cloud Storage recovered almost four times from an IT- related downtime event than organizations that stored business critical data on-site. It is also noted that companies that utilized public Cloud Storage had a superior disaster recovery program than those that did not do so. In terms of RTO goals, the study said that although the RTOs for clouds and non-cloud storage users were the same about 12 hours, users of Cloud storage, met this RTO goal 100% of the time while those who failed to adopt Cloud storage were able to meet only 80% of the objective.

While all these findings, put up a strong case for Cloud based storage, what was particularly heartening was the fact that majority of the early adopters identified disaster recovery as the primary driver for using Cloud storage services. This essentially means that these early buyers of Cloud storage have their eyes firmly set on the end objective i.e. an efficient enterprise-wide DR strategy and, if their experience with cloud storage are anything to go by, the day might not be far off when they might move other pieces of their DR set up in the public cloud.

REFERENCES

1) Cass, S (2009), Market watch: Virtual computers, real money, MIT/Technology Review, July/August http://www.technologyreview.com/computing/22608/
2) Chopra Anil, " Cloud Computing: For Enterprises or SMBs", PC Quest, October.
3) Cloud security.org http://cloudsecurity.org
4) Google suggest http://www.google.com/webhp?complete=1&hl=en
5) How stuff works.com http://communication.howstuffworks.com/
6) Albert Greenberg, James Hamilton, David A. Maltz, and Parveen Patel. Cost of a cloud: Research problems in data center networks. In ACM SIGCOMM Computer Communications Review, Feb 2009.
7) Kimberley Keeton, Cipriano Santos, Dirk Beyer, Jeffrey Chase, and John Wilkes. Designing for Disasters. Conference On File And Storage Technologies, 2004.