Comparative analysis of regular and cloud disaster and recovery systems

Jasmin Azemović
Faculty of information technologies
University Džemal Bijedić
Mostar, Sjeverni logor bb, Bosnia and Herzegovina
jasmin@edu.fit.ba

Elmin Sudić
Faculty of information technologies
University Džemal Bijedić
Mostar, Sjeverni logor bb, Bosnia and Herzegovina
elmin.sudic@edu.fit.ba

Abstract. *Cloud infrastructure is very popular among companies that have large business requirements that based on information technology infrastructure. One of the major problems is relatively high price of investment which requires their data centers for continuous work without investments to recovery systems. Reducing risk of data loss costs are relatively high, hence majority of small and midsize businesses decide not to take significant steps for constructing disaster and recovery systems (DR). This work discusses existing solutions and mechanisms to provide continuous work and fast recovery for low tolerance systems. Moreover, this work discusses the cost-effectiveness of utilizing cloud systems as a DR measure against private hot site DR system.*

Keywords. disaster, recovery, cloud

1 Introduction

Most of today's business is based on computer systems and information technologies. If those systems fail they cost companies lot of money and is some cases it can threaten human lives. It is necessary to invest a lot of effort and money to reduce down time because continuous work is key factor for business success [1]. There are two types of IT users, one who never lost data and one who make backup. It is estimated that only 6% of companies suffering from a catastrophic data loss can survive, while 43% never reopen and 51% close within two years [2]. Lose data because error on hard drive is just one of problems, errors and mistakes also can include loss of whole data centers.

According to the study by Whitehouse and Buffington in [3], hardware errors lead to 45% of all failures, human errors 31%, software failures in 14% of the infections companies, errors due to the virus are 7% and system failures due natural disasters are 3%. National Computer Security Association (NCSA) shows that 50% of critical data for companies is stored at unsafe locations like desktop or laptop computer of employees and not encrypted [4]. Only 34% of companies that make backup to secure data of lost test created backup and recovery procedures [3]. Looking at the aforementioned statistical facts cloud computing could assist in overcoming DR problems. Building a fast recovery system is challenging and has relatively high costs. Therefore introducing cloud systems could assist in decreasing the costs and provide faster disaster recovery. Moreover virtualization on cloud systems makes it easier to create and use DR which is that is more reliable and cost-effective than regular solutions [5].

The rest of the paper is organized as follows: Section II discusses business continuity (BC); Section III discusses DR mechanisms; Section IV discusses about startup difference between standard and cloud DR systems. Section V discusses cloud based recovery solutions; Section VI overviews prices of cloud DR services; Section VII covers the pros and cons of cloud computing. Finally in Section VIII we give conclusion of the paper and discuss future work.

2 Business continuity

Business continuity (**BC**) is an activity which allows companies to secure critical business data and make them available for employees and customers [6]. Business continuity is not an activity which is implemented in time of catastrophe because BC covers all daily activities, service maintenance and disaster and recovery. In BC, there are seven tiers of disaster and recovery. It was originally defined by Share to help identify the various methods of recovering mission-critical computer systems as required to support business continuity. The seven tiers of business continuity solutions offer a simple method to define current service levels and associated risks. Successful disaster and recovery depends on two time Recovery Point Objective (RPO) and Recovery Time Objective (RTO) [1, 6].

Recovery Point Objective is the maximum acceptable amount of data loss measured in time. It is the age of the files or data in backup storage required to resume normal operations if a computer system or network failure occurs [6]. RPO time depends on the importance of data, some applications must not lose any data (RPO = 0).

Recovery Time Objective is the maximum desired length of time allowed between an unexpected failure or disaster and the resumption of normal operations and service levels. The RTO defines the point in time after a failure or disaster at which the consequences of the interruption become unacceptable [1, 6].

RTO include error detection, server preparation (virtual and physical), installation application and establishment of network communications to enable redirection from original to backup site [1]. Figure 1, shows recovery time and prices of tiers depending on business continuity plan:

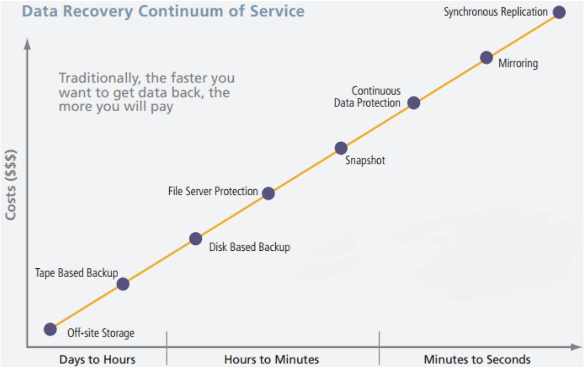


Figure. 1. Recovery time and prices depending on business continuity plan[1]

where the lowest tier has highest recovery time, however it is relatively cheap and the highest tier has the shortest recovery time, yet it is relatively expensive. The following sections discuss about recovery procedures for each tier.

2.1 Tier 0: No off-site data

Businesses with a tier 0 business continuity solution have no Business Continuity Plan. Information is not saved, no documentation, no backup hardware, and no contingency plan. The time necessary to recover in this instance is unpredictable if not impossible to recover at all [7].

2.2 Tier 1: Data backup with no hot site

Businesses that utilize tier 1 continuity solutions back up their data and send them to an off-site storage facility. The method of transporting these backups is often referred to as Pick-up Truck Access Method (PTAM). Depending on how often backups are created and shipped, these organizations must be prepared to accept several days to weeks of data loss, but their backups are secure off-site. However, this tier lacks the systems on which to restore data [7].

2.3 Tier 2: Data backup with hot site

Businesses utilizing Tier 2 business continuity solutions make regular backups on tape. This is combined with an off-site facility and infrastructure known as a hot site in which to restore systems from those tapes in the event of a disaster. This solution will still result in the need to recreate several hours or even days’ worth of data, but the recovery time is more predictable [7].

2.4 Tier 3: Electronic vaulting

Tier 3 solutions are built on the components of tier 2. Additionally, some mission critical data is electronically vaulted. Data is constantly copied over safe communication canals to backup server. This electronically vaulted data is typically more current than that which is shipped via PTAM. As a result there is less data recreation or loss after a disaster occurs [7].

2.5 Tier 4: Point-in-time copies

Tier 4 solutions are utilized by businesses that require both greater data currency and faster recovery than users of lower tiers. Rather than relying largely on shipping tape, as is common on the lower tiers; tier 4 solutions begin to incorporate more disk based solutions. Several hours of data loss is still possible, but it is easier to make such point-in-time (PiT) copies

[1] Figure 1. Image courtesy of Network World

with greater frequency than tape backups even when electronically vaulted [7].

2.6 Tier 5: Transaction integrity

Tier 5 solutions are used by businesses with a requirement for consistency of data between the production and recovery data centers. There is little to no data loss in such solutions, however, the presence of this functionality is entirely dependent on the application in use [7].

2.7 Tier 6: Zero or near-Zero data loss

Tier 6 business continuity solutions maintain the highest levels of data currency. They are used by businesses with little or no tolerance for data loss and who need to restore data to applications rapidly. These solutions have no dependence on the applications or applications staffs to provide data consistency. Tier 6 solutions require some form of disk mirroring. There are various synchronous and asynchronous solutions available from the mainframe storage vendors. Each solution is somewhat different, offering different capabilities and providing different recovery point and recovery time objectives. Often some form of automated tape solution is also required. However, this can vary somewhat depending on the amount and type of data residing on tape [7].

2.8 Tier 7: Highly automated, business integrated solution

Tier 7 solutions include all the major components being used in a tier 6 solution with the additional integration of automation. This allows a tier 7 solution to ensure consistency of data above that which is granted by tier 6 solutions. Additionally, recovery of the applications is automated, allowing for restoration of systems and applications much faster and more reliably than would be possible through manual business continuity procedures [7].

3 DR mechanisms

Traditionally work with disaster and recovery systems requires access to physical resources and secured copies of data which are stored in remote locations. There are different types of strategy or mechanisms for storing data on remote sever, but three of them are most typical: cold, warm and hot site. This thermal description is equal how much time it takes to recover and how much it cost disaster and recovery. Cold site is cheapest solution but disaster recovery time is longest. Hot site is the most expensive disaster and recovery solution but it ensures fastest recovery time when disaster occurs. Warm site is solution in the middle between cold and hot site in terms of cost and disaster recovery time [1].

3.1 Cold site

A cold site is a type of backup site to which a business or government agency can go if the primary site is destroyed, damaged, or otherwise rendered inoperable.

Cold sites are generally the least expensive type of backup site to establish, but they are also somewhat less effective than more expensive options that can be operational within hours of a primary site going offline.

A cold site is a backup site that does not necessary need to have hardware and equipment already in for operational use but it can have it. Cold site network infrastructure need to be configured and customized like primary site including VLAN, VPN, DNS and firewall with same roles. Post disaster primary site needs to be restored on cold site in period of hours or days depending on the cold site infrastructure. Lot of organizations won’t spend money on testing cold site solution and consequently disaster recovery is done by principle of trial and error. Very often they can’t manage to restore data and system functionalities [1].

3.2 Warm site

Standby **Warm site** implies maintaining copies of data on hard drives, and very often companies create virtual machines to speed up disaster recovery time. Organization that use warm site is able to recover data in range from minutes to hours. Testing hard drives in some period of time is much easier then maintenance of tapes. Warm site is often used like replication or mirroring backup server and disaster and recovery system. Virtual machines include operating system, applications, business data and preconfigured settings which facilitate synchronization between primary and warm site [1,9].

3.3 Hot site

Standby **Hot site** provide fastest recovery system after disaster, because data and application is up to date and running. This type of recovery site enables constant application availability even in situation when primary site is down. Multiple instances of the application can be set to do update and one of them can provide service to users. This type of disaster and recovery solution can provide instant redirection to hot site when error occurs and provide same services as primary site. Cluster of servers and synchronized replication are good DR options, but they are most expensive and hardest to maintain.

There is lot of research about price of hardware and software designed for disaster and recovery systems. Research is showing that these systems are showing research that less than 20% medium-sized businesses (100 to 9999 employees) have in the corporate property secondary data storage [1,8].

Same research shows that 48% medium-size business and 23 % big companies rent storage places for disaster and recovery infrastructure [8]. Maintaining second site that is corporate property or it is rent cost companies lot of money, and losses are relatively large. Costs include buying and maintenance of servers, hard discs and network infrastructure including site renting of place, cooling systems, employees' salaries and energy bills.

4 Startup time difference between regular and cloud recovery system

Based on the type of business and importance of data it is necessary to define the type of disaster recovery systems. Standard backups are done with magnetic tape, CDs, HDD etc. which needed to be stored in special environmental conditions on farther geographical distance from primary site [2]. Backup storage on a further distance requires more time to store and restore the data. Cloud DR system enables data mirroring and instant switch to recovery site [10]. Based on these facts we see that time needed to start recovery site on cloud is measured in seconds or minutes, while standard recovery systems need several hours or days [7].

5 Cloud like disaster and recovery solution

Cloud computing based on virtualization present different approach for implementation disaster and recovery plans. Based on virtualization all servers, operating systems, applications and data are encapsulated in one package [5, 10]. Whole virtual server can be copied as one package and stored or utilized as virtual server, since virtual servers are hardware independent.

A consequence of a virtual server being encapsulated in one package; it can be copied and safely transferred from one to another data center without moving each component of server. This approach can significantly reduce time of disaster recovery compared to standard DR solution. With virtualization systems can be restored in minutes or in seconds if it is replication server [10].

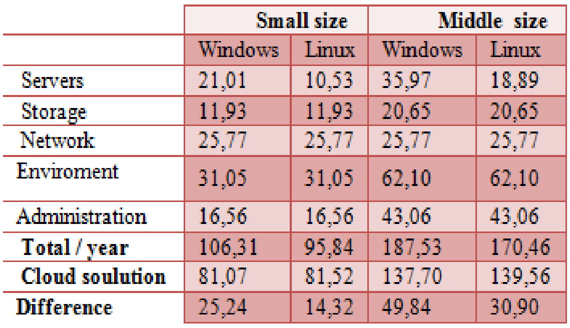
Cloud computing disaster and recovery becomes more cost-efficient with much faster recovery time. Besides a fast recovery after failure, cloud computing offers more options. Leading providers offers high availability for services which are installed on their servers. In order to make system scalable, safe and available like cloud systems companies will have to invest lot of money, but lot of companies do not have that amount of money to spent for DR solutions [8].

According to research by Amazon, 51% of companies that use cloud like DR solution which utilize replication and virtual machines enable them to decrease RTO or to speed up disaster recovery. ESF research states that 43% of users consider that it is easier to do DR tests for 45% of large companies and 32% for small companies [8].

6 Price of cloud DR service

Considering cloud computing possibilities like DR solution we see advantages comparing to the standard DR approach (fast recovery, performance, easy deploy etc.). High availability implies and great investment for organizations that want fastest disaster recovery. Fastest recovery solution is implemented with hot site however this solution is the most expensive. In order to show costs of these systems we compare cost of private and cloud hot site DR solution. Comparison is done with two types of organizations, small (up to 50 employees) and medium size organization (up to 250 employees) based on two platforms Windows and Linux. The amount of IT equipment is defined by the type of the company's business and number of employees. We supposed that small organizations use several servers (web, mail, DNS server and some monitoring services etc.), while lager business requires more complex infrastructure with more IT equipment. It should be emphasized that we observed whole system infrastructure which includes hard drive space, network infrastructure, firewalls and internet provider services. Cost comparison is done by Amazon

Obtained prices with Cost Comparison Calculator [11] are compared with prices of some other cloud DR solutions such as Microsoft, Google and VMware etc. so we for final calculation use average price from all providers is calculated. Some of providers have larger prices for their services while others have lower with same or very similar performance. In order to make final report we calculate one year of using cloud and private hot site DR solution.



	Small size		Middle size	
	Windows	Linux	Windows	Linux
Servers	21,01	10,53	35,97	18,89
Storage	11,93	11,93	20,65	20,65
Network	25,77	25,77	25,77	25,77
Enviroment	31,05	31,05	62,10	62,10
Administration	16,56	16,56	43,06	43,06
Total / year	106,31	95,84	187,53	170,46
Cloud soulution	81,07	81,52	137,70	139,56
Difference	25,24	14,32	49,84	30,90

Table 1. Exact costs of calculations (000 USD)

Table 1. shows price of all components included in system architecture. Cost of single part of system is sum of prices for all individual components that are included. All prices that are used in calculation are very similar for all cloud providers so there are no significant differences between them.

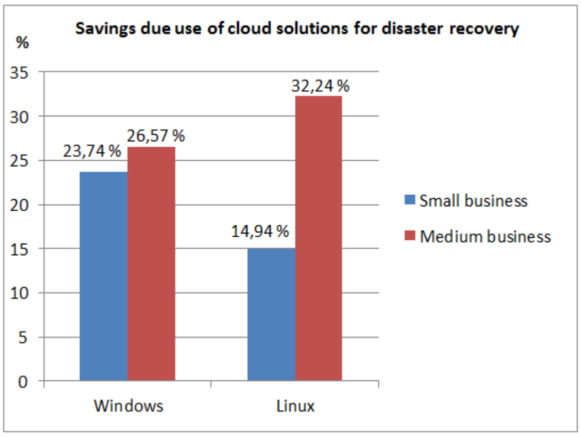


Figure. 2. Savings due use of cloud solutions for disaster recovery

Starting hot site on Windows platform small companies need to pay 23.74% more in order to use private DR infrastructure. In case of middle business for companies which utilize this platform and private DR infrastructure need to pay 26.57% more money for three years of use. Starting services on Linux platform requires less investment than Windows because of product license and need for higher performance hardware. Small companies in there years will save 14.94% of money with cloud instead of private infrastructure. Middle size organization will save 32.24% money with cloud DR solution instead of same solution on private infrastructure.

7 Pros and cons of cloud computing

Cloud computing has some pros and cons. However depending on the type of business some disadvantages are acceptable, while of some advantage don't have great benefit. For example news sites don't have privacy problems because all data is public and also data ownership after use of cloud in not the question.

DR mechanism needs to adapt to architecture change very fast in order to provide BC. Cloud systems are ideal for reconfiguration when it is needed. Migrations and cloning enable and simplify failback and it is useful planned maintenance downtime [1]. But cloud systems have some security concerns that are still not solved and that is it major disadvantage.

In order to decide to use cloud or not as a DR solution or to completely move the business to cloud depends on how the pros will outweigh cons as in the following table 2. Even there is no clear line about cloud being a good solution or not, it is important to consider all facts that make it unique solution.

Pros	Cons
Cost reduction	Privacy
Scalability and Performance	Data mobility and ownership
Easier collaboration	Security in the Cloud
Almost Unlimited Storage	Prone to Attack
Easy Access to Information	Dependency and vendor lock-in
Quick Deployment	Technical Difficulties and Downtime
Latest version availability	Limited control and flexibility
Device Diversity and Location Independence	High bandwidth
Smaller learning curve	Not always more cost-effective
Resiliency and Redundancy	Specific applications

Table 2. Cloud computing pros and cons

According to Miller in [12], there are several types of businesses for which cloud computing is considered to be a good solution: collaborators, road warriors, cost-conscious users, users with increasing needs. Additionally to the user that the cloud is a good solution gives examples of customers that the cloud is a bad solution: the internet-impaired, offline workers, the security-conscious.

8 Conclusion

Taking into consideration advantages of what cloud computing offers it is clearly that this is technology that will continue to grow. Nowadays all big companies and vendors like Google, Microsoft, Amazon etc. offers some of the cloud services (mail, online storage etc.). Lot of these and similar companies have plans for future to completely shift their business in cloud [8]. Currently there are no standards for cloud computing. However IBM start to work on „The Open Cloud Manifesto“ [13] which is relatively large step forward to standardize cloud computing. Cloud computing is getting more popular fans because it is easy to use and offers pay-as-you-go model [1]. In order to use cloud computing users don’t need to buy hardware, operating systems and network infrastructure return on investment is very fast. A key factors which slows the spreading and popularization of cloud computing is because users have fear about privacy, losing control of data and systems and because they already spend money for private infrastructure. As a consequence, a lot of organizations won’t leave their existing infrastructure; 43% of users say that they will use their resources in next five years [8].

However lot of organizations and private customers already use some type of cloud service. Shifting their business to cloud will allow organizations to have cheap start which enables bigger competitiveness in business. Another huge thing is that vendors every day do updates and upgrades that make cloud computing even more accessible and easier to use.

References

[1] Emmanuel Cecchet, K.K. Ramakrishnany, Prashant Shenoy, Jacobus van der Merwey, and Arun Venkataramani, *Disaster Recovery as a Cloud Service: Economic Benefits & Deployment Challenges Timothy Wood*.

[2] Jim Hoffer, *Backing Up Business - Industry Trend or Event*, *Health Management Technology*, Jan 2001.

[3]http://www.sounditservices.com/services/offsite-disaster-recovery/ 4.10.2013.

[4] Michael Miora, *Enterprise Disaster Recovery Plan (NCSA)*, 1996.

[5] Cam Macdone and Paul Lu, *High-Performance Computing: A Quantitative Study of Basic Overhead.*

[6] Prof. Kurt J. Engemann, *International Journal of Business Continuity and Risk Management*, 2011.

[7] Charlotte Brooks, Matthew Bedernjak, Igor Juran, John Merryman, *Disaster Recovery Strategies with Tivoli Storage Management*, November 2002.

[8] Lauren Whitehouse and Jason Buffington, *Amazon Web Services: Enabling Cost-Efficient Disaster Recovery Leveraging Cloud Infrastructure*, Jan 2012.

[9] *Disaster Recovery Best Practices Guide Riverbed Technical Marketing*, October (2011)

[10] Al Muller, Seburn Wilson, *Virtualization with VMware ESX Server*, 2005

[11] http://aws.amazon.com/tco-calculator/ 6.10.2013.

[12] Michael Miller, *Web-Based Applications That Change the Way You Work and Collaborate Online*, 2008.

[13] http://www.opencloudmanifesto.org/ 7.10.2013.