

Assessing Disaster Recovery Alternatives: On-site, Colocation or Cloud

Omar H. Alhazmi, Ph. D.,
Taibah University
Dept. of Computer Science
Medina, Saudi Arabia
ohhazmi@taibahu.edu.sa

Yashwant K. Malaiya, Ph. D.,
Colorado State University
Dept. of Computer Science
Fort Collins, CO, USA
malaiya@cs.colostate.edu

Abstract— Every organization requires a business continuity plan (BCP) or disaster recovery plan (DRP) which falls within the cost constraints while achieving the target recovery requirement's in terms of recovery time objective (RTO) and recovery point objective (RPO). The organizations must identify the likely events that can cause disasters and evaluate their impact. They need to set the objectives clearly, evaluate feasible DRPs to choose the one that would be optimal. Here we examine tradeoffs involved in choosing among the disaster recovery options. The optimal disaster recovery planning should take into consideration the key parameters including the initial cost, the cost of data transfers, and the cost of data storage. To evaluate the risk, the types of disaster (natural or human-caused) need to be identified along with the probability of a disaster occurrence and the costs of corresponding failures needs to be evaluated. An appropriate approach for the cost evaluation needs to be determined to allow a quantitative assessment of currently active disaster recovery plans (DRP) in terms of the time need to restore the service (associated with RTO) and possible loss of data (associated with RPO).

Keywords- *Cloud, Disaster Recovery, RPO, RTO, Risk Analysis and Management*

I. INTRODUCTION

Business continuity is a vital requirement of many businesses, as a sudden service disruption can directly impact business objectives causing significant losses in terms of revenue, business reputation and losses of market share. Indeed, some organizations may find it difficult to survive a serious disaster [1]. The causes of disasters can either be unintended events such as a power failure or intentional such as a denial of service attack. Consequently, an organization must have a disaster recovery plan (DRP) which is executable, testable, scalable and maintainable. Such a plan must satisfy cost constraints while achieving the target recovery objectives; that is, recovery time objective (RTO) and recovery point objective (RPO) [2]. The organizations involved must identify likely events that can cause disasters and evaluate their impact. They need to set the objectives clearly, and evaluate feasible disaster recovery plans to choose the optimal DRP.

Many smaller organizations may find it difficult to afford a desirable disaster recovery plans. Hence, some may choose to have only periodic data backups. This is due to the fact that traditional disaster recovery plans often depend on having two identical sites: a primary and a secondary site, which may be located at some distance. Unfortunately, having two sites will add significantly to IT cost for a

disaster that is likely to occur only rarely and therefore may seem like unjustified overhead. This may explain why around 40-50% of small businesses have no DRP and no current future plans to have one [3].

Fortunately, the cloud computing technology that has emerged recently which provide an affordable alternative to traditional DRPs for small or medium sized businesses, with minimal startup cost and with no significant addition to staffing and office space costs [4,5]. Public cloud services generally use “pay-for-what-is-used” model which can make the secondary site on the cloud very cost effective. The cost is divided among the many users of public cloud services, who may actually use these services only occasionally.

II. BACKGROUND

A key concept in a DRP is the physical separation of the primary and backup sites. A significant fraction of disasters including those caused by outages are geographical [6].

When active processing of incoming transactions is switched from the failed primary to the backup site, the switch is termed a **failover**. When the causes of the primary failure have been addressed and the switch is made back to the primary, the switch is termed a **fallback**.

A number of options arise depending on the nature of the backup site and how it links to the process at the primary site. The backup site is often described as follows.

Cold standby: Recovery in such a case requires hardware, operating system and application installation. Thus recovery can take multiple days

Hot standby: This requires a second data center that can provide availability within seconds or minutes. A hot site can take over processing while the primary site is down. A complete copy of the primary process may sometimes exist at the backup, with no need to install either the OS or the application.

Warm standby: A tradeoff between a hot and a cold site.

It should be noted that the terms “hot” and “warm” are sometimes defined differently. Recovery levels are sometimes described in terms of tiers [7]. The tiers are characterized by two key measures, recovery time objective (RTO) and recovery point objective (RPO) which are the main objectives that need to be satisfied criteria when evaluating the optimal solution with a given overall cost. They are described below.

RTO: The duration in which business functions is unavailable and must be restored (includes time before disaster is declared and time to perform tasks). RTO depends on the tasks needed to restore the transaction handling

capabilities at the backup server. While it can take days using tape backups, it can take less than a minute in advanced systems.

RPO: The duration between two successive backups, and thus the maximum amount of data that can be lost when restoration is successful. Historically the maximum value has been 24 hours. If the backup is a synchronous mirrored system, RPO is effectively zero.

III. EVALUATION OF DRP SCHEMES

Here we examine the factors that need to be considered to evaluate the system cost, assuming that the year is used as the period for computing costs. The total annual system cost C_T is the sum of the initial cost C_i (amortized annually), ongoing cost C_o plus the expected annual cost of potential disasters C_d

$$C_T = C_i + C_o + C_d \quad (1)$$

The ongoing cost C_o is the sum of ongoing storage cost C_{os} , data transfer cost C_{ot} , and processing cost C_{op} :

$$C_o = C_{os} + C_{ot} + C_{op} \quad (2)$$

The annual disaster cost is the total expected cost of disaster recoveries plus the cost of unrecoverable disasters. For a disaster type i , let the probability of disaster occurrence be p_i , and the let two costs be C_{ri} and C_{ui} . Then

$$C_d = \sum_i p_i (C_{ri} + C_{ui}) \quad (3)$$

Note that the recovery cost includes the cost of using the backup after the failover and the cost of lost transactions. The cost of lost transactions is proportional to the RTO duration. Loss of reputation is another factor to consider.

Some disaster frequency related data is available but needs to be analyzed to develop a model. The geographical correlation factor needs to be modeled to determine potential statistical correlation between primary and backup failures. When disasters are rare, using a shared cloud may reduce the cost significantly since expensive cloud services will only be needed when a disaster strikes.

RPO is the time between two successive backups. It is an implementation dependent variable. Its optimal value would depend on the overhead represented by a data backup [8, 9]; however, it may be determined based on scheme used.

RTO determines the length of the period during which the system is not available for incoming transactions. It depends on the factors that impact the DRP tier level used. Let the delays for the backup be as follows.

T1 = hardware set-up/initiation time

T2 = OS initiation time

T3 = Application initiation time

T4 = data/process state restoration time

T5 = readiness verification time + IP switching time

RTO would depend mainly on the readiness of the backup site. At the minimum, it would include T5. For a site that starts out completely cold, T1 to T5 would be required.

$$RTO = \text{fraction of RPO} + \sum_{j=\min}^5 T_j \quad (4)$$

Where j min depends on the service readiness of the backup. The fraction of RPO represents computation lost

since the last backup. At this point in time there is not enough data to construct completely analytical models to determine the optimal implementation. However, this paper can serve as a guide to evaluate available alternatives.

Some cloud service providers provide calculators or pricing guides that permit estimation of costs. Examples of such computations now exist in the literature [2].

Several feasible alternatives should be identified and evaluated. Using these schemes we shall examine three different DRP backup options:

Onsite: the backup and the system both in one location,

Co-location: the backup site is remotely located.

Cloud: the backup site is located in the cloud.

Looking at these options we can analyze them using the given equations. It is clear that for each option these variables vary significantly; thus, significantly impacting the cost. Hence, the decision maker can be unbiased to any of these options when doing the feasibility study and business requirements analysis by using reliable quantitative metrics.

IV. CONCLUSIONS AND FUTURE WORK

Development of analytical methods can guide future planning and maintenance of a DRP. A quantitative approach will allow CIOs to compare applicable DRP solutions to select and specify an optimal one. Besides, there is a need to collect more data to permit the development of models that can eventually allow the problem to be set up as a mathematical optimization. These include relationship between geographical distance and statistical correlation between failures in the primary and secondary sites. A model relating RTO and cost can be developed. Some of the literature speculates that there may be a non-linear relationship between cost and RTO [7], this needs to be further investigated.

REFERENCES

- [1] Disaster Recovery for Small Business, Technical White Paper, Iomega Corporation, March 18, 2009.
- [2] T. Wood, E Cecchet, K. K. Ramakrishnan, P. Shenoy, J. van der Merwe, and A. Venkataramani, "Disaster recovery as a cloud service: economic benefits & deployment challenges", Proc. 2nd USENIX conference on Hot topics in cloud computing (HotCloud'10), Berkeley, CA, USA, 2010, pp. 8-8.
- [3] Survey Indicates Half of SMBs Have No Disaster Recovery Plan, Chris Preimesberger, September 2009, <http://www.eweek.com/c/a/Data-Storage/Survey-Indicates-Half-of-SMBs-Have-No-Disaster-Recovery-Plan-687524>.
- [4] M. Pokharel, S. Lee, J. S. Park, "Disaster Recovery for System Architecture Using Cloud Computing", IEEE/IPSJ Int. Symp. Applications and the Internet, 2010, pp. 304-307.
- [5] M. Wiboonratr and K.i. Kosavitsutte, "Optimal strategic decision for disaster recovery," Int. Journal of Management Science and Engineering Management, Vol. 4 (2009) No. 4, pp. 260-269.
- [6] [Symantec 2010 Disaster Recovery Study, Global results](#), CA, USA, November 2010.
- [7] Disaster Recovery Strategies with Tivoli Storage Management, C. Brooks, M. Bedernjak, I. Juran, J. Merryman, IBM/Redbooks, November 2002.
- [8] Disaster Recover/Business Continuity, N-1 Technologies, http://www.n-1technologies.com/recovery_continuity.html.
- [9] The Total Cost of (Non) Ownership of a NoSQL Database Cloud Service, Jinesh Varia and Jose Papo, March 2012, http://media.amazonwebservices.com/AWS_TCO_DynamoDB.