A SWOT ANALYSIS ON CISCO® HIGH AVAILABILITY VIRTUALIZATION CLUSTERS DISASTER RECOVERY PLAN

Eman Al-Harbi
431920472@student.ksa.edu.sa

Soha S. Zaghloul
smekki@ksu.edu.sa

Faculty of Computer and Information Science
Department of Computer Science
King Saud University
Riyadh, Saudi Arabia

Abstract

Continuity of services availability is critical in some business types such as banks, hospitals, and military institutions. However, this may be interrupted by either a natural disaster such as an earthquake, volcano, or a tornado. In addition, human negative interference such as hackers or terrorism may expose valuable resources to danger. Moreover, disasters may occur because of a network interruption, system crash, or electricity outage. Therefore, a disaster recovery plan is vital to provide the precautions necessary to minimize the negative effects that may occur as a result of a potential disaster. Many companies in the market provided solutions for disaster recovery to resume operations after a failure with the minimum losses. This paper studies the Cisco ® disaster recovery plan, and then performs a SWOT analysis on the provided solution.

Keywords Disaster Recovery Plan; Cisco; SWOT analysis; Warm standby; Hot standby; Distributed Data Center;

1. Introduction

Cloud computing involves the delivery of services over the Internet to geographically dispersed hosts. Many services are offered through the cloud such as software (SaaS), platform (PaaS) and Infrastructure (IaaS). A data center is the physical location that accommodates the dense arrangements of computer equipment, associate networking, telecommunications, storage and auxiliary equipment required to store, process, manage and disseminate data and information [1]. One of the most useful benefits of cloud computing is the rapid provision of a flexible, scalable, and cost-effective data center. Therefore, the disaster recovery plan (DRP) of a data center should be highly considered for two main reasons. The first reason is that the probability of a system crash is directly proportional to the workload exerted on the resources; this should be under focus especially if the data center belongs to one of the sensitive tenants such as a bank or a hospital. The second reason is that the recovery plan in fact shifts the cost of the secondary facility to the budget of another party. According to [2], "*Data Center Disaster Recovery is the organizational planning to resume business operations following an unexpected event which may damage or destroy data, software and hardware systems*".

In our life, many types of disasters may be encountered. Some of them are natural disasters such as flood, earthquake, tornadoes, etc… Other disasters may occur because of the interference of criminal human-beings, such as hackers or even terrorists, who intentionally target the loss of resources. In addition, operational disasters may occur; examples are loss of network connections, power failure, data corruption or loss, false redundancy, or cascading systems failure. The danger of such disasters lies in the services interruption rather than the severity of the disaster per se. Imagine that the services of a bank are interrupted for just two seconds: this entails the loss of customers' money and may cost the bank a

wealth. Another example is the operation of the medical systems: the interruption of such type of services may cost humans' lives.

A disaster recovery plan should ensure the availability of the application all the time. In case of a disastrous event, all data should be automatically and transparently directed to an alternate site. DRP solutions may be categorized into three types: (1) Corrective: in which the solution is activated after the occurrence of a disaster to rectify the situation; (2) Detective: the aim of such plan is to discover undesirable situations; and (3) Preventive: measurements are taken to avoid the occurrence of any undesirable event [3].

This paper studies a data center disaster recovery plan (DRP) developed by Cisco ® that uses the High Availability (HA) virtualization clusters. The rest of the paper is organized as follows: Section 2 explains the data center disaster recovery metrics. Section 3 exposes the DRP approaches. Section 4 explains briefly Cisco's solution for DRP. In Section 5, a SWOT analysis is performed on Cisco's solution. Finally, Section 6 concludes the paper.

2. DATA CENTER DISASTER RECOVERY METRICS

In order to develop a DRP, a study is conducted to identify the possible types of disasters that may impact the business under consideration. For each type of failure, an estimation of the possible losses is made. The length of time spanned between the service failure and its recovery is critical.

Juniper® uses two metrics to measure the ability to restore or to keep continuous service [4]; these are:

2.1 Recovery Point Objective (RPO)

This is an indication of how much data loss the system is able to tolerate. Therefore, for ultimately sensitive data, this amount should be at its minimum. On the other hand, the RPO can be large enough if the data loss within a specific period of time is tolerable. In other words, the RPO is directly proportional with the allowable amount of lost data. Shorter RPOs are achieved by continuous data protection techniques.

2.2 Recovery Time Objective (RTO)

The second metric, the RTO, is an indication of how much time is needed to achieve data recovery. It is measured in units of time that may range from minutes to days. The value of the RTO is directly proportional with the level of criticality of the data under consideration.

Methods to accomplish high values of RPO/RTO include high availability (HA), clustering, virtualization clustering, and data replication. Worth to mention that there is a correlation between the RPO/RTO from one side and the network on the other side. In other words, in order to achieve a zero data loss RPO with very low RTO, then a high-bandwidth, low-latency network is required. This requirement becomes even more critical if a large amount of data is involved.

3. DATA CENTER DRP APPROACHES

Three typical DRP redundancy strategies are found in the market. Here they are:

3.1 Cold Standby

In cold standby, the data of the primary server is mirrored to the secondary server on a regular basis, as in warm standby. However, mirroring is performed less frequently as compared to the warm standby. In case of the failure of the primary server, the backup server is called upon. This

approach is used for non-critical data, or in case there are infrequent data changes.

The cold standby is simple to implement and easy to maintain. However, it suffers from a substantial delay to move from the standby mode to the active mode. RPO and RTO are high in this scenario: RPO may extend to several weeks, whereas the RTO may reach several hours.

3.2 Warm Standby

In this approach, two servers are used: the primary and the secondary. The former contains the original data. The secondary captures data from the primary server at regular intervals of time, without causing its interruption. Accordingly, the data on both servers is not the same all the time.

Warm standby necessitates the manual intervention in case of a disaster. Both RPO and RTO are considered reasonable as compared to the cold standby.

3.3 Hot Standby

In this strategy, two identical servers are used. In other words, the data is sent simultaneously to both the primary and the backup servers. Therefore, data is identical in both servers all the time. This redundancy strategy is used in case of critical data. However, it consumes from the network bandwidth.

Data mirroring is accomplished in real-time in this scenario. Therefore, RTO is measured in minutes whereas the RPO approaches zero.

4. CISCO DRP SOLUTION

Cisco devised a disaster recovery (DR) and a disaster avoidance (DA) solution based on the distributed virtual data center strategy [5]. The reported framework provides solutions to all resources including processors, storage and network. The main idea relies on moving the virtual machines (VMs) installed on the physical data center to another one. This is independent of the separating distance between the two data centers. However, it depends on the available resources at the destination.

Cisco offers multiple network services in order to optimize the access and the distribution to the remote sites. Here they are:

- An intelligent domain name system (DNS) to redirect the requests from end-users to the active physical location. It also contributes in the load balancing across multiple active sites.
- HTTP traffic redirection between sites: in case of insufficient resources, the end-user is automatically and transparently redirected to a more suitable backup data center, where information and resources are available.
- Route health injection: provides a real-time, granular distribution across multiple sites based on application availability.

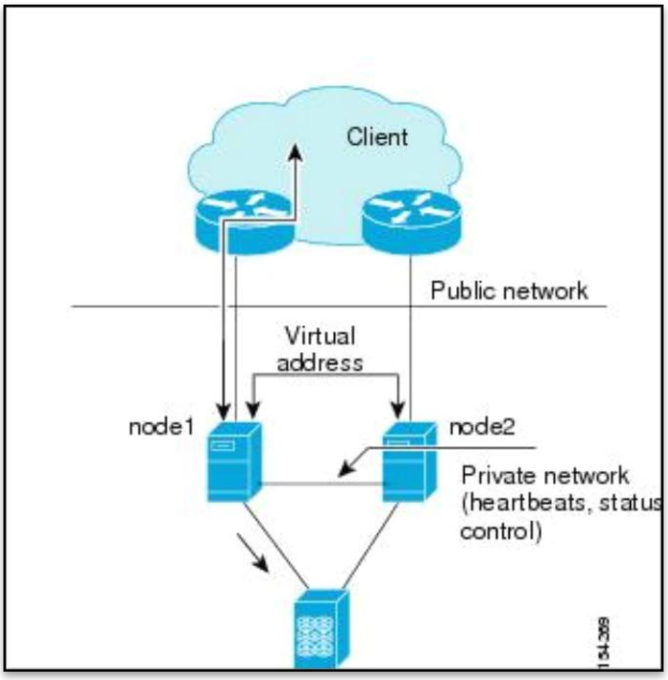

Figure 1: High Availability Clusters Design

Cisco's solution use server high availability clustering; therefore, multiple copies of applications are available to allow access to data read and/or write. In fact, clustering provides availability, reliability, scalability, and manageability [6]. Figure 1 [7] illustrates such design.

Cisco's solution uses geoclusters: these are high availability clusters that span a wide geographic area. Examples include campus, metro, regional and continental clusters. However, geoclusters face many challenges such as split brain, layer 2 hearbeat, and storage disk zoning [6]. In split brain, multiple active nodes access the same disk simultaneously, which results in data corruption. A Quorum, a tie breaker, is therefore used to resolve such contention. Layer 2 (L2) heartbeat results of the extension of L2 network. In addition, the extension of virtual LANs (VLANs) is hazardous. This is resolved by configuring an Ethernet over MPLS (EoMPLS). MPLS stands for Multi-Protocol Label Switching: data are directed from one node to the next based on short path labels instead of long network addresses. This avoids the complexity of looking into a routing table [8]. When a failure occurs, it takes over from a disk array: this is known as storage disk zoning. In order to avoid this problem, the cluster enabler instructs the disk array to perform a failover in case of a failure event. Table 1 summarizes the challenges faced by a HA cluster, and their solutions.

Table 1. HA Cluster Challenges and Solutions

Challenge	Solution
Split brain	Quorum
L2 Heartbeat	EoMPLS
Storage Disk Zoning	Failover

Finally, Cisco uses two modes of data replication: synchronous and asynchronous. In synchronous mode, all data written to local and remote arrays is completed and acknowledged before an I/O instruction. On the other hand, in the asynchronous mode, the writes are acknowledged and the I/O command is complete on the local disk array only. Changes are replicated to remote arrays asynchronously. Table 2 highlights the difference between the two modes.

Table 2. Comparing Synchronization Modes

Item	Synchronous	Asynchronous
Application Performance	High	No impact
Distance	Limited	Unlimited
Data Loss	No data loss	Possible

5. SWOT ANALYSIS

In this section, an analysis is performed on the previously presented Cisco solution. Points of strength and weaknesses are highlighted. In addition, opportunities and threats are listed.

5.1 Strengths

Cisco solution makes use of both HA clusters and virtualization technology. Accordingly, it captures the benefits of both of them. HA clusters provide high availability of data and application. In addition, it provides easy manageability. Moreover, data centers are flexibly located with no restrictions on distances.

Furthermore, the virtualization technology facilitates live migration for the sake load balancing; therefore, the possibilities of failures are decreased. Furthermore, virtualization provides better resources utilization. Therefore, capital, operational, management and maintenance costs are reduced.

Moreover, the usage of disk arrays provides increased availability, maintainability, and resiliency by using additional hardware redundantly such as controllers and power supply.

In addition, synchronization duplication mode ensures data integrity. Asynchronous mode has no impact on the application performance and has no restrictions on the distance between the data centers.

5.2 Weaknesses

Naturally, data mirroring takes initially a very long time when issuing a backup server. On the other hand, hot standby may be considered expensive for non-critical data. The HA clusters suffer from split brain which may cause data corruption. In addition, L2 hearbeat limits the networks extension. Moreover, storage disk zoning affects the disk array bandwidth.

The success of the Cisco solution depends to a far extent on the network bandwidth and latency. Thus, the implementation must ensure the network is reliable enough.

Furthermore, synchronous replication mode negatively affects the application performance and restricts distances lengths. On the other hand, the asynchronous replication mode can cause loss of data: this is not tolerable in case of sensitive information.

5.3 Opportunities

Clusters provide a good solution for scalability; therefore, multiple sites may be created and networked together whenever needed.

Asynchronous mode has no limitations on the distance between data centers; therefore, the DRP may extend to additional distant sites.

In fact, choosing distant sites located in different continents decrease the possibilities of the harms caused by natural disasters. In case a flood, for example, takes place in a continent, DRP may be activated and sensitive data accordingly migrate to another continent.

EoMPLS may be correctly configured to guarantee distant transfer of data without involving the complexities of routing.

Locally, disk arrays may be added as much as needed.

Finally, the hot standby, if used, guarantees the minimum RTO and RPO.

5.4 Threats

In order to make the optimum use of the solution, live migration policy aspects should be carefully taken of. In addition, the disaster recovery plan should continuously be considered for improvement and adaptation to the environment changes.

Furthermore, a decision may be taken for one or more VMs to migrate, however, the hardware resources are not sufficient at the destination. On the other hand, VMs are vulnerable during live migration; they may be easily attacked. Security is a critical concern in live migration.

The DRP is threatened in case of the simultaneous migration of a huge amount of data over the network. For example, if a natural disaster happened somewhere, the DRP should be activated to all involved sites. This may cause bottlenecks in the network, high latency, or even may lead to the whole network failure in the extreme case.

Table 3 summarizes the SWOT analysis.

6. CONCLUSION

Natural, human-induced and operational disasters may cause tremendous loss for businesses. Data corruption or loss, revenue loss, and customers' dissatisfaction may result from such disasters. Therefore, businesses invest a lot of money in DRP. Many leading companies developed efficient DRPs with different strategies. This paper focuses

on Cisco HA Virtualization Clusters DRP. It was found that there is a trade-off between the DRP approaches: a comprehensive study of the level of data criticality should be made to save money and provide safety. The plan should estimate the target RPO and RTP, and the solutions are accordingly provided. On the other hand, HA clusters face many challenges such as split brain, L2 heartbeat, and storage disk zoning. Furthermore, there is a tradeoff between the synchronous and asynchronous mode. Plan designers should study what factors can be sacrificed of application performance, long distances and data integrity, before deciding on which mode to deploy.

Table 3. SWOT Analysis Summary

Strengths	Weaknesses
1. High availability of data and applications 2. Easy manageability 3. Flexibility in locations and environment 4. Ensures load balancing 5. Provides better resources utilization 6. Reduced costs 7. Increased manageability and resiliency 8. Data integrity for synchronous duplication mode 9. No impact for asynchronous duplication mode	1. Initial mirroring consumes a lot of time 2. Expensive in case of hot standby for non-critical data 3. Subject to data corruption 4. Limited networks extension 5. Negative impact on disk array bandwidth in case of a failure 6. Dependency on network bandwidth and latency 7. Synchronous duplication mode negatively affects the application performance 8. Synchronous duplication mode restricts distances 9. Loss of data in case of asynchronous duplication mode
Opportunities	Threats
1. Scalability 2. No limit on distances in case of synchronous duplication mode 3. Diversity in data center locations allows more precautions to natural disasters 4. EoMPLS allows the network extensibility 5. Duplication of disk arrays as much as needed locally 6. Minimum RTO and RPO with hot standby mode	1. Insufficient hardware resources. 2. Inefficient live migration policy aspects 3. Inadaptability to new amounts of data and new environment 4. Vulnerability of VMs during live migration 5. Insufficient network bandwidth in case of an overwhelming disaster

7. REFERENCES

[1] Integral Group. “Energy Efficiency Baselines for Data Centers”. Whitepaper, December 2012.

[2] Core Link Data Center, http://www.corelink.com/glossary/data-center-disaster-recovery.

[3] http://en.wikipedia.org/wiki/Disaster_recovery.

[4] Juniper Networks. “MPLS Data Center Interconnection for Disaster Recovery”. Whitepaper, 2011.

[5] Cisco Networks. “Distributed Virtual Data Center for Enterprise and Service Provider Cloud”. Whitepaper, 2012.

[6] KwaiSeng Consulting Engineers. “Data Center Disaster Recovery”. Whitepaper, Cisco Networks, 2007.

[7] Cisco Networks. “Data Center High Availability Clusters”. Whitepaper.

[8] http://en.wikipedia.org/wiki/IP/MPLS.