REMOTE MIRRORING: A DISASTER RECOVERY TECHNIQUE IN CLOUD COMPUTING

Ashish Bajpai[1], Poonam Rana[2], Seema Maitrey[3]

[1,2,3] Department of Computer Science & Engineering,
Krishna Institute of Engineering and Technology, Ghaziabad (India).

ABSTRACT

As we know Cloud Computing offer many advantages such as scalability, multi-tenancy of resources and cut hardware costs using virtualization technologies like Xen, Hypervisor software . VMware is also a brand name in virtualization technology. But what happens when a catastrophic failure occurred to the whole cloud storage system Until now there is no such provision to secure our private and confidential data. We have three services models available in the market PaaS(Platform as a service), SaaS(Software as a service) and IaaS(Infrastructure as a service) . In this paper we proposed another cloud service model disaster recovery as a service and a architecture how we can increase the fault tolerant of cloud service model during catastrophic failure. So we use a technique called remote mirroring to protect our data using ISCSI initiator.

Keywords: Cloud storage system, DRaaS(Disaster recovery as a service), Remote mirroring , ISCSI initiator.

I INTRODUCTION

Cloud computing is an internet based computing, where we can share our resources, software and information, are provided to computers and devices on demand, it is an emerging model for business computing[1]. The cloud computing is the combination of traditional computing technology and network technology like parallel computing, distributed computing and so on. Cloud computing systems aim to data calculation and procession[2].The cloud computing can be divided into two models service models and deployment models deployment models contains Public cloud, Private cloud and Hybrid cloud and the service cloud contains the IaaS, PaaS,SaaS. The aim of cloud computing is to construct a perfect system with powerful computing capability and of low cost by making use of the advanced business model like SaaS(Software as a Service), PaaS(Platform as a Service), IaaS(Infrastructure as a Service) to distribute the powerful computing capacity to it's users. Cloud storage is a new concept extended and developed from the concept of cloud computing[3] In this paper we are proposing a new service that is Disaster Recovery as a Service(DRaaS) which provide technical support during catastrophic failures.

II RELATED WORK

Zhang JianHua and Zhang Nan explains the Cloud Computing based Data Storage and Disaster Recovery in that paper they suggest the idea to solve the problem of database service such as storage capacity, performance, stability, security, and many other issues, cloud storage was used to provide cloud based data platform and also they describe the architecture of cloud and presents the deployment of the disaster recovery and other applications in inter-private cloud storage helps in achieving true cloud computing. In this paper they first explain the cloud storage system in which they talk on the different layers of the storage system and those layers are the Storage layer, Infrastructure Management layer, Application Interface layer, Access layer. And after that they talk on the Disaster Recovery of Cloud Storage in which they propose an architecture of Disaster Recovery system and they also proposed a model of Disaster Recovery System.[4]

Chao-Tung Yang and Alfredo Cuzzocrea talks on Improvement of Cloud Virtual Machine Availability with Virtualization Fault Tolerance Mechanism. They suggest that a virtualization is a common strategy to enhance the pre-occurring computing resources, particularly in the field of cloud computing. One of the Apache projects popularly known as Hadoop is specifically designed to scale up from single servers to thousand of machines where each offer local computation and storage. In this paper they paper they try to reached Hadoop high availability which called Virtualization Fault Tolerance (VTF). The Apache project firstly explained as : Hadoop and then they explain High availability, after that Fault Tolerance Technology, followed by Virtualization Technologies h along with Dynamic Resource Allocation. [5]

III PROBLEM STATEMENT

There are various problems existing in cloud storage system, the cloud storage system poses new challenges in data security, reliability, and management and there is no provision of disaster recovery in the existing cloud system and in this paper we have proposed a scheme to save the cloud storage system from disaster by giving a new term fault tolerance as a service [FaaS] and by creating a second site point on a remote location where we can store our data and can provide the best and trustable services to the user, and also the user need not to worry about the data security

3.1 Cloud disaster recovery

While applications and data services are visible, wicked attacks must be avoided. It is important and difficult to guarantee data security in cloud storage system. It is necessary to create fault tolerant function in less cost for cloud storage. The fault tolerant function can overcome single point failure and avoid data loss. At the same time, there must be fault-tolerant backup system, which can ensure that data have a reliable backup if they are lost due to catastrophic failure.

In order to satisfied the continuity of application and the security of data, the structure of disaster recovery system is "distributed computing, centralized storage". According to different necessities, disaster recovery has three levels. They are data-level disaster recovery, system- level disaster recovery, and application-level disaster recovery. Data-level disaster recovery is the most basic and it can ensure the security of the application data. System- level disaster recovery disaster has further requests for operating system of application server, making disaster recovery time can be as short as possible. System-level disaster recovery requests real-time by which users could not feel that any disaster occurred.[6]

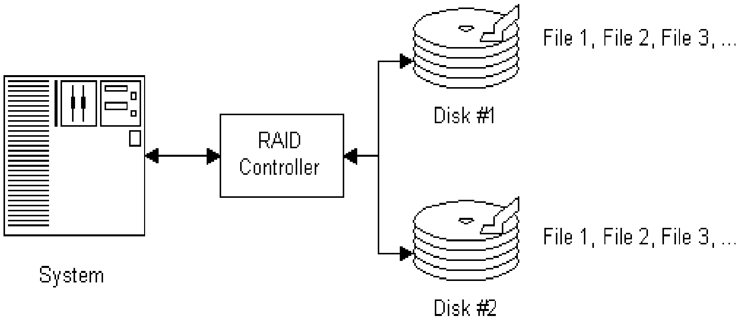


Fig 1: Remote Mirroring Technique

3.2 Remote mirroring as a disaster recovery technique for cloud computing

Remote mirroring is a classic technique for tolerating failures: by keeping two or more copies of important information, access can continue if one of them is lost or becomes unreachable. It is used inside disk arrays (where it is called RAID1), between disks or disk arrays, and across multiple sites, where it is called *remote mirroring*.[7]

Many technologies are involved in building the disaster recovery system, e.g. SAN/NAS technology, remote mirroring technology, IP-based interconnect technology, snapshot, etc. But here we talk on remote mirroring technology.

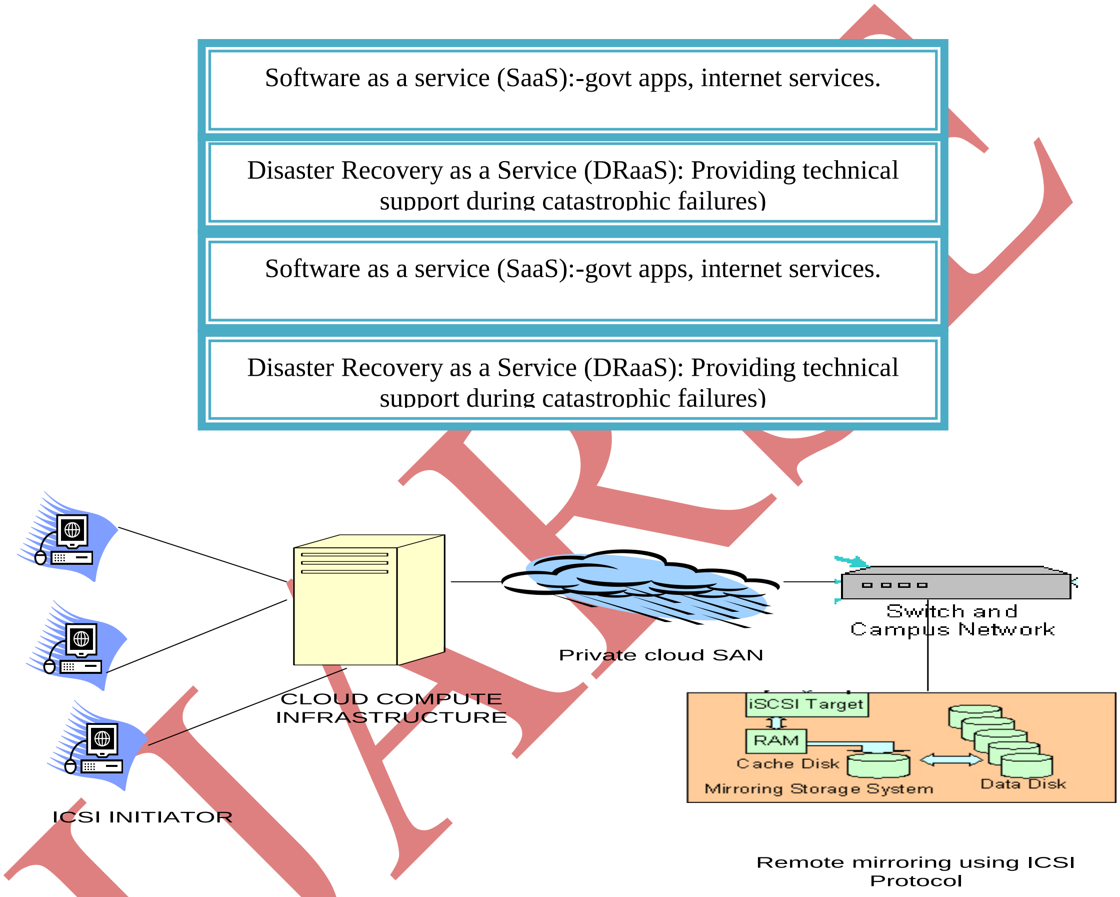
3.3 Remote mirroring technology

Remote mirroring technology is described in Fig-1 used during the data backup between the primary data center and backup center. Mirroring is the information storage process to create mirror views for the same data on two or more disks or disk subsystems, one primary mirror system and one secondary mirror system. According to the location of the primary/secondary mirror storage system, it can be classified into local mirror and remote mirror. Remote mirror, also called remote replication, is the core technology of disaster recovery and the basis to keep remote data synchronization and realize disaster recovery. According to whether the host requesting mirror needs confirmation information from the remote mirror site, the remote mirror is classified into synchronous remote mirror and asynchronous remote mirror.

Synchronous remote mirror (synchronous replication technology) means to replicate local data to a remote location in a fully synchronized way through the remote mirror software; each local I/O transaction shall be released after receiving the confirmation for completion of remote replication. Synchronous mirror enables remote copy to always match the replicated content requested by the local machine. In case of any failure with the primary site, the mirrored remote copy can ensure continuous business execution without any data loss after the user's application switches to the alternative site. However, due to the long round-trip transmission delay, it only applies to relatively short distance.

Asynchronous remote mirror (asynchronous replication technology) ensures the completion of basic I/O operations on the local storage system before updating the remote storage view, and the local storage system will provide confirmation information for the completion of I/O operations to the host requesting mirror. Remote data replication is made synchronously in the background, which minimizes the impact on system performance, ensures long transmission distance (up to 1,000 kilometers), and requires low network bandwidth. However, as the write of many remote dependent storage subsystems is not confirmed, there may be problem of data inconsistency if data transmission fails due to certain factors. To solve this problem, replication delay technology is mostly adopted (local data replication is made in the background log area) to update data remotely while ensuring local data completenes

3.4 Remote Mirroring Architecture



Software as a service (SaaS):-govt apps, internet services.

Disaster Recovery as a Service (DRaaS): Providing technical support during catastrophic failures)

Software as a service (SaaS):-govt apps, internet services.

Disaster Recovery as a Service (DRaaS): Providing technical support during catastrophic failures)

Switch and Campus Network
Private cloud SAN
CLOUD COMPUTE INFRASTRUCTURE
iSCSI Target
RAM
Cache Disk
Mirroring Storage System
Data Disk
ICSI INITIATOR
Remote mirroring using ICSI Protocol

Fig 2: Architecture of Remote Mirroring In Cloud Computing

3.5 Cloud Disaster Recovery: Five key Steps to Avoid Risk and Protect Your Data

Introducing applications on the cloud is appealing to IT organizations for many reasons. First, they benefit from the markets of scale and buying power that a large data center can gather. Second, cloud providers generally offer hardened data centers, backup power sources and other capabilities that only large organizations can afford. Most important, IT organizations can take advantage of these services on a pay-as-you-go basis or guaranteed handiness, depending upon organizational requirements. Many organizations today do not have adequate disaster recovery (DR) protection for their applications. In most cases, disaster recovery is perceived as too expensive, complex.[7]

Because many hosting providers maintain multiple data centers, IT managers often assume that disaster recovery (DR) is either intrinsic in the architecture or that DR is not an issue that warrants anxiety. However, DR attentiveness is not the default configuration for most providers that offer cloud storage infrastructure. One multi-national media firm for which Acquainted checked is surprised to learn that its cloud-hosted key applications had no DR coverage whatsoever. The 9/11 attacks taught us many things about IT disaster preparedness, including the fact that epic disasters are extremely rare but not unimaginable.

IT managers who have hosted applications through cloud providers, or are thinking of doing so, should perform the same DR due-diligence they would for in-house infrastructure. This includes assessing the risks, laying out the potential solutions and implementing a plan that meets the required service level at the least cost. In most cases, providers will accommodate requested DR configurations but will not have their own reference architecture. IT managers should, therefore, be prepared to work with their cloud providers to help architect and specify an appropriate solution. IT organizations requiring assistance with this process can partner with us on any of the steps, from assessment to planning, architecting, specifying and implementing cloud DR solutions. This paper discusses the storage elements of disaster recovery planning.

Step 1: Assessing the Risks

Disaster risks can be thought of as a continuum of probability, but they can also be categorized in three major groups:

Site disaster: Fire, broken pipes or long-term power outages that can render the data center (or computer components) unusable for longer than the specified service agreement.

Area disaster: Floods, tornados, hurricanes, major snow storms and even pandemics can render a data center unusable.

Regional disaster: Terrorist attacks, financial failures, train derailments with toxic chemicals, pandemics, etc.

Proper data center design mitigates many of the risks associated with all three of these categories. Data center location can mitigates likely weather events, train derailments and the like. Hardened data centers will have battery power backup (uninterruptable power supply) for sudden power loss and power generators for longer term power needs (which could extend for as much as a few days).

Even so, an area or regional disaster can render a data center unusable even if it continues to fully exist physically. One of the most likely risks that would require a disaster response is the financial failure of a hosting service. When the “tech bubble” burst early in the last decade, many hosting providers of the late 1990’s failed financially, leaving customers to scramble for infra- structure alternatives. While such an event does not carry the immediacy of a tornado response, organizations that have a plan in place to re-host applications will experience less disruption and likely lower costs than those forced to scramble at the last minute. Obviously, a DR plan to move your applications to another site of the same hosting company would not be viable for such a contingency.

The quaint notion of IT personnel boarding a plane with backup tapes in hand, headed for a recovery site is a thing of the past. There is no certainty that these sites will be available nor that transportation will be operating. As part of the risk assessment, IT managers must consider how they would complete a recovery if transportation were severely restricted or if their chosen site could not be accessed.

Step 2: Determining Requirements

After completing an assessment of the risks, IT organizations need to classify their recovery requirements for the applications that are hosted. Requirements should be developed in the context of:

Recovery Point Objective (RPO): RPO is the degree to which data loss can be tolerated. For example, "immediate RPO" indicates zero- tolerance for data loss. A 24-hour RPO would indicate that restoring data as of yesterday's backup is sufficient, resulting in a loss of all transactions and data after that time.

Recovery Time Objective (RTO): RTO determines the maximum tolerable time for recovering the data and bringing the application back online. Note: The RTO should include the time to restart systems, databases and applications and re-route communications; simply restoring the data alone is not enough.

Both RPO and RTO requirements are driven by the cost of downtime. Cost of downtime can include actual loss of revenue, loss of employee productivity, loss of customer goodwill and loss of reputation. Tangible financial losses are the easiest to consider and most directly correlate to the cost of mitigation. If the cost of downtime is greater than the cost of a specific DR strategy, then the strategy is obviously worth the cost. Loss of customer goodwill and reputation are less tangible, but just as important. If the cost of acquiring and keeping a customer is high, then a stringent DR plan may be worth it. In any event, it is a business decision that involves weighing risks and consequences against costs.

Step 3: Understanding DR

A typical DR service works by replicating application state between two data centers; if the primary data center becomes unavailable, then the backup site can take over and will activate a new copy of the application using the most recently replicated data[8].The most basic element of a DR plan is getting the data outside the data center. But how far outside? If the secure facility is less than 10 miles from the data center, the distance is insufficient to guard against area disasters and pandemics. It is possible that neither the data center nor the secure storage facility would be accessible at the same time. A common guideline calls for 90 miles of separation between locations to guard against such events as "dirty" bombs. IT organizations can decide for they how far is far enough. Data can be stored offsite asynchronously on tape or disk, or synchronously (disk only).

Backup to tape and offsite storage: Tape remains the cheapest method for moving data to a second site or archiving it. There are some "gotchas" IT managers should consider with tape backup:

Tape format: The format between the source and the target must be compatible. For example, LTO tapes cannot be read by DLT tape drives. Beware of different generations of the same technology as well.

Backup and recovery software: These software applications write the data in proprietary formats (unless explicitly specified to write in TAR or CPIO formats). For example, CommVault Galaxy cannot restore a tape created by Symantec Net Backup.

Asynchronous site-to-site data replication: This method moves data offsite but to magnet- ic disk drives vs.

magnetic tape. Disk backups can significantly reduce recovery time in the event of a disaster.

Most backup/recovery applications can back up to disk using compression and/or de-duplication technology. Thus, the backup image is much smaller than the actual data image. With change-only backup methods, transmission bandwidth and data storage requirements are minimized.

Virtual tape libraries (VTLs) are specialized storage devices (disk-only or disk-to-tape) that can further automate the DR process.

Synchronous site-to-site data replication: Synchronous data replication ensures that every piece of data entered or changed is simultaneously replicated. While synchronous replication is typically the most expensive off-site replication option, for some critical applications the cost may be justified. Synchronization delivers an immediate RPO and an RTO limited to the time that it takes to declare a disaster, restart the application from the second site and re-establish communication.

Synchronous DR does not eliminate the need for a backup and recovery solution. Even where the cost of synchronous data replication is justified, synchronous backup will typically be combined with asynchronous methods for point- in-time restore capabilities supporting rollback to previous application or file versions. Whatever the backup methodology, it is important to ensure that the system image (i.e., the operating system) be included with the backup set so that the entire environment can be recreated if necessary.

DR Mechanisms

Disaster Recovery is primarily a form of long distance state replication combined with the ability to start up applications at the backup site after a failure is detected. The amount and type of state that is sent to the backup site can vary depending on the application's needs. State replication can be done at one of these layers: (i) within an application, (ii) per disk or within a file system, or (iii) for the full system context. Replication at the application layer can be the most optimized, only transferring

the crucial state of a specific application. For example, some high-end database systems replicate state by transferring only the database transaction logs, which can be more efficient than sending the full state modified by each query [9]. Backup mechanisms operating at the file system or disk layer replicate all or a portion of the file system tree to the remote site without requiring specific application knowledge [10].

Step 4: Auditing Cloud Providers

Cloud providers should be willing to provide users with documentation regarding their data center protection strategies and, in fact, many have published literature describing these features. IT managers should compare this against their own list of requirements, just as they would for their own data center. They should examine it for location and should not assume that the hosting company considered nearby rail facilities or manufacturing operations that use toxic chemicals when choosing the site.

Elements that should be considered in a cloud provider audit include:

- Location
- Possible events

- Power grid/communications considerations and contingencies
- Proximity to potential terrorist targets (e.g., airports, seaports, national landmarks)
- Relationship to recovery destinations
- Data center hardening features
- Vendor’s DR contingencies

Step 5: Implementing and Managing Your Cloud DR Solution

Over time, your organization will evolve. Applications will be added, enhanced or abandoned; offices will open and close; new technologies and formats will emerge. Moving forward, many cloud providers will likely be on the buying or selling side of a data center acquisition and integration. As important as it is to evaluate and select the right provider and the right DR solution, it is equally important to review your DR requirements and solutions over time. Make sure your cloud provider has instituted a process for simulating and testing your DR solution and ensuring that all systems are performing as promised. A rolling quarterly test of a subset of applications may be sufficient, as long as most or all of your systems are eventually tested on an annual basis. Internal requirements should also be reviewed

Proper Planning Prevents Poor Performance

IT managers should not passively assume that their cloud hosting provider has the disaster recovery contingency addressed. In most cases, it is a capability that must be explicitly requested. Moreover, IT managers should be prepared to drive the conversation and deliver specifications to the provider. Most providers are accommodating but often lack the expertise to guide customers toward an appropriate solution.

Just as with an in-house solution, IT managers should assess the potential risk of disasters and the impact of a protracted recovery. Any solution must consider the necessary RPO and RTO required by the application. It is not necessary to have the ultimate recovery for every application, so IT managers should balance the cost of providing a specific RPO/RTO against the cost of downtime for the application[11].

IV ADVANTAGES AND LIMITATIONS OF REMOTE MIRRORING FOR DR

For an application using physically separate locations, the Remote mirroring provides data accessibility safeguard. Remote mirroring takes place over MAN or WAN distances as similar to mirroring within a RAID array. It's usually aligns between storage arrays or storage appliances, and can be synchronous or asynchronous.

The highest possible level for DR recovery point objective (RPO) and recovery time objective (RTO) is managed by Synchronous remote mirroring. The RPO is "zero" lost data, and the RTO is typically seconds to minutes. Synchronous remote mirroring does this by neither completing nor acknowledging the local write until the remote write is completed and acknowledged. Additional writes can't occur until each preceding write has been completed and acknowledged. This means local performance is directly related to the performance of the DR remote device; distance is the limiting factor. Remote synchronous mirroring is rarely deployed or circuit distances greater than 160km (100 miles).

With asynchronous remote mirroring, local writes are completed and acknowledged before the remote writes.

Asynchronous remote mirroring is a "store-and-forward" technique that reduces I/Os and wait delays, allowing remote writes to fall behind the local writes. This means the RPO for lost data can range from seconds to minutes, and even hours in some cases. Asynchronous remote mirroring is most often utilized when the remote site is a long distance from the local site.

The primary advantage of both synchronous and asynchronous remote mirroring is the minimal (asynchronous) to zero (synchronous) risk exposure in losing data during a disaster. A secondary advantage is the potential for quick data recovery when a disaster occurs. Remote mirroring doesn't require server agents, and it provides heterogeneous server and application support.

Remote mirroring applications are often pricey, the equipment is usually expensive, and it typically requires at least twice the primary disk space and sometimes much more. However, when the lowest possible RPO and RTO are the requirement, remote mirroring is the answer.

Another disadvantage is that remote mirroring doesn't prevent a rolling disaster, data damage, corruption or accidental deletion. If data is corrupted, damaged or deleted at the primary site, it will also be at the DR site. Some asynchronous remote mirroring products timestamp each transaction and allow recovery to a point in time before the corruption or deletion occurred, but they're exceptions to the rule. This means procedures other than remote mirroring must also be implemented to allow for recovery of corrupted, damaged or deleted data. Other disadvantages include lack of support for heterogeneous arrays, no support for internal storage, and nearly no application and file information.

Less-expensive alternatives to remote mirroring can also provide the lowest possible RPO and RTO. They're generally continuous data protection (CDP) products and include time-based continuous snapshots, automated backup, replication of changed data and automated, generational-change distributed backup. They offer a lower TCO than remote mirroring, support heterogeneous storage and provide better rollback capabilities. But they usually require installing and managing agents.[12]

V CONCLUSION AND FUTURE WORK

In this paper we have proposed an architecture for our cloud storage system during catastrophic failure i.e. Disaster recovery as a service (DRaaS) using remote mirroring technique. But this is a costly and expensive technique to replicate data using remote mirroring so this is the overhead can be incurred in implementing this technique as a framework. So the future work is that what we can do to minimize this overhead.

REFERENCES

[1] ARMBRUST M, FOX A, GRIFFITH R, et al. “*Above the Clouds: A Berkeley View of Cloud Computing*” [R]. Berkeley, CA,USA: University of California, 2009.

[2] Dai Yuanshun. “*The Brief Review of Cloud Computing Technologies*”. Information and Communications Technologies. 2010.2,pp 29-35.

[3] ZHOU Ke, WANG Hua, and LI Chunhua. “*Cloud Storage Technology and Its Application*”, ZTE Communications, 2010.16(4),pp 24-27.

[4] Zhang jian hua, and Zhang Nan, "*CloudComputing-based data storage and disaster recovery*, IEEE 2011 International Confrence on Future Computer Science and Education.

[5] Chao-Tung Yang , AlfredoCuzzocrea, "*On Improvement of Cloud Virtual Machine Availability with Virtualization Fault Tolerance Mechanism*" IEEE 2011. 2011 Third IEEE International Conference on Cloud Computing Technology and Science pg no.122-129.

[6] QU Ming-cheng, WU Xiang-hu,; LIAO Ming-hong, et al. "*A Disaster- Tolerant Storage Model and a Low Data Failure Model for Data Grid*". Acta Electronica Sinica. 2010.38(2),pp 315-320.

[7] Seneca: remote mirroring done write Minwen Ji, Alistair Veitch, John Wilkes HP Laboratories, Palo Alto, CA

[8] Timothy Wood, Emmanuel Cecchet, K.K. Ramakrishnany,Prashant Shenoy, Jacobus van der Merwey, and Arun Venkataramani . "*Disaster Recovery asCloudService:Economic Benefits & Deployment Challenges*". In Proceeding HotCloud'10 Proceedings of the 2nd USENIX conference on Hot topics in cloud computing Pages 8-8.

[9] Tirthankar Lahiri, Amit Ganesh, Ron Weiss, and Ashok Joshi. Fast Start:quick fault recovery in oracle. ACM SIGMOD Record, 30(2), 2001

[10] Kimberley Keeton, Cipriano Santos, Dirk Beyer, Jeffrey Chase, and John Wilkes. Designing for Disasters. Conference On File And Storage Technologies, 2004.

[11] Vmware high availability. http://www.vmware.com/products/high-availability/.

[12] Phil Goodwin, "Cloud Disaster Recovery: Five key Steps to Avoid Risk and Protect Your Data", Cognizant 20-20 insights, 2011.