Effectiveness of Backup and Disaster Recovery in Cloud
A Comparative study on Disk and Cloud based Backup and Disaster Recovery

Akash Kaveti
This thesis is submitted to the Faculty of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Masters in Electrical Engineering with emphasis on Telecommunication Systems. The thesis is equivalent to 20 weeks of full time studies.

**Contact Information:**
Author: 
Akash Kaveti 
E-mail: akka14@student.bth.se

External advisor:  
Sujoy Chatterjee  
General Manager- IT Services  
ITC Infotech India Ltd.

University advisor:  
Adrian Popescu  
Department of Communication Systems  
School of Computing  
BTH, Karlskrona

Faculty of Computing  
Blekinge Institute of Technology  
SE-371 79 Karlskrona, Sweden  
Internet : www.bth.se  
Phone : +46 455 38 50 00  
Fax : +46 455 38 50 57
**ABSTRACT**

**Context:** Definitive information is what makes enterprises in the contemporary world continuously move up the ladder in today’s world. Hence they need to be always up to date with the latest advancements. A key point in the enterprise structure is Backup and Disaster Recovery, DR system and this has very high scope for development since it can be integrated with a lot of our daily requirements. On account of this, it has a lot of applicability in future for development of technology. Due to this, we have chosen a Fast Moving Consumer Goods Company, FMCG to conduct experimentation on their Backup and DR system.

**Objectives:** In every organization, Backup and DR plays a crucial role in Business Continuity Planning. This work relates to associate backup and recovery plan with an organizational view. And thereby work on its association with Recovery Time Objective, Recovery Point Objective, time taken for backup, time take for recovery and Total Cost of Ownership.

**Methods:** Literature study is the first step to understand present scenario of trending technologies. Hence our understanding led us to conduct an experimental setup where we compared data that was collected in a case study and evaluated its performance of Backup and DR problems that are faced in today’s technology dominated world.

**Results:** In our research, we took various parameters into consideration, which affect performance of an enterprise Backup and DR system. This drove us to assess disk-based and cloud-based Backup and DR plans in the FMCG environment.

**Conclusions:** In conclusion, we summate that even though there have been a lot of theories and research on how Backup and DR is vital, we still lag behind in research work on which the organizations can rely upon to shift towards advanced technologies without risking their competence. With sufficient research on the entrepreneurial environment, we can improve performance enterprises and improve our present knowledge about cloud Backup and DR thereby improving its conventional usage in the present world.

**Keywords:** Backup, Disaster Recovery, Performance, RTO and RPO
ACKNOWLEDGEMENT

Immeasurable appreciation and deepest gratitude for the help and support are extended to the following persons who in one way or another have contributed in making this study possible.

Prof. Adrian Popescu, Supervisor who was abundantly supportive, advices, guidance, provision and valuable suggestions which have helped me in completing this thesis study. I gratefully acknowledge the support and guidance of Prof. Kurt Tutschku for his valuable time and efforts to help us during the course.

I thank Hardeep Garewal for providing us an opportunity to conduct this thesis study in an enterprise and gaining a knowledgeable experience. I would like to thank Sujay Chatterjee for instilling the characteristics of a good researcher in me. Discussions during this industrial thesis with Sujay Chatterjee are invaluable to me. Bhujay Kumar Bhatta for his enthusiasm and unlimited zeal in training our view on this thesis. Also, Gan Sharma for his constructive and valuable inputs for this study.

Many co-workers have helped me in conducting and documenting this thesis. I would like to thank Karthik Krishna, Mohan Kumar, Emmanuel Anthony, Amit Kumar Pandey, Pushpalatha Kuruba, Jagadish Kb for their innumerable suggestions which were most valued. And I would like to thank Anoop Kodakara, Latheesh Devairakkam, Panna Lal Shaw, and Sanket Srivastav for providing us with equipment required.

I would like to thank the survey respondents for their valuable time in answering the survey and providing me with inputs.

Last but not least, I would like to thank my parents and sister for providing me with the ability to complete this program. The constant support they provide during thick and thin keeps motivating me to perform better, every day.
<table>
<thead>
<tr>
<th>ABBREVIATIONS</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR</td>
<td>Disaster Recovery</td>
</tr>
<tr>
<td>JBOD</td>
<td>Just a bunch of Disks</td>
</tr>
<tr>
<td>AWS</td>
<td>Amazon Web Services</td>
</tr>
<tr>
<td>S3</td>
<td>Simple storage service</td>
</tr>
<tr>
<td>DS</td>
<td>Data System</td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>NAS</td>
<td>Network-attached Storage</td>
</tr>
<tr>
<td>OS</td>
<td>Operating System</td>
</tr>
<tr>
<td>RPO</td>
<td>Recovery Point Objective</td>
</tr>
<tr>
<td>RTO</td>
<td>Recovery Time Objective</td>
</tr>
<tr>
<td>SAN</td>
<td>Storage Area Network</td>
</tr>
<tr>
<td>TCO</td>
<td>Total Cost of Ownership</td>
</tr>
<tr>
<td>VM</td>
<td>Virtual Machine</td>
</tr>
<tr>
<td>SCSI</td>
<td>Small Computer System Interface</td>
</tr>
<tr>
<td>SATA</td>
<td>Serial AT attachment</td>
</tr>
<tr>
<td>WAN</td>
<td>Wide Area Network</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS

ABSTRACT ........................................................................................................................... I
ACKNOWLEDGEMENT ........................................................................................................... II
ABBREVIATIONS .................................................................................................................. III
TABLE OF CONTENTS ........................................................................................................ IV
LIST OF FIGURES ................................................................................................................. V

1 INTRODUCTION ................................................................................................................. 6
  1.1 PROBLEM STATEMENT ................................................................................................. 7
  1.2 OBJECTIVES ................................................................................................................. 7
  1.3 RESEARCH QUESTIONS ............................................................................................... 7
  1.4 OUTLINE OF THE THESIS ......................................................................................... 8

2 RELATED WORK AND BACKGROUND ...................................................................... 9
  2.1 FULL BACKUP ............................................................................................................. 9
  2.2 INCREMENTAL BACKUP .......................................................................................... 10
  2.3 FULL + INCREMENTAL BACKUP ........................................................................... 10
  2.4 DIFFERENTIAL BACKUP ......................................................................................... 11
  2.5 FULL + DIFFERENTIAL BACKUP .......................................................................... 11
  2.6 SYNTHETIC BACKUP ........................................................................................... 12
  2.7 FILE LEVEL STORAGE .......................................................................................... 12
  2.8 BLOCK LEVEL STORAGE ..................................................................................... 12
  2.9 VM LEVEL BACKUP ............................................................................................... 12
  2.10 JBOD ...................................................................................................................... 12
  2.11 AWS DIRECT CONNECT ...................................................................................... 13
  2.12 PERFORMANCE METRICS .................................................................................. 13
     2.12.1 RTO and RPO .............................................................................................. 13

3 METHOD .......................................................................................................................... 15
  3.1 LITERATURE REVIEW ............................................................................................. 15
  3.2 SURVEY ................................................................................................................... 16
     3.2.1 Participants of survey .................................................................................... 16
  3.3 CASE STUDY ........................................................................................................... 16
     3.3.1 Backup and DR using Disk ........................................................................... 17
     3.3.2 Backup and DR using cloud .......................................................................... 18

4 RESULTS AND ANALYSIS ......................................................................................... 19
  4.1 SURVEY RESULTS .................................................................................................. 19
  4.2 EXPERIMENTATION RESULTS .............................................................................. 20
  4.3 ANALYSIS ............................................................................................................... 25

5 CONCLUSION .................................................................................................................. 26

REFERENCES ...................................................................................................................... 28

APPENDIX A ....................................................................................................................... 30
APPENDIX B ....................................................................................................................... 37
LIST OF FIGURES

FIGURE 1, FULL BACK UP[13] .................................................................................................................. 9
FIGURE 2, FULL AND INCREMENTAL BACKUP[13] .............................................................................. 10
FIGURE 3, FULL AND DIFFERENTIAL BACKUP[13] ............................................................................. 11
FIGURE 4, SATA/SAS JBOD[21] ........................................................................................................... 13
FIGURE 5, RPO AND RTO[23] ............................................................................................................. 14
FIGURE 6, RESEARCH METHODOLOGY OVERVIEW ......................................................................... 15
FIGURE 7, BLOCK DIAGRAM OF DISK BACKUP .................................................................................. 17
FIGURE 8, CLOUD BACKUP AND DR BLOCK DIAGRAM ................................................................. 18
FIGURE 9, VM INCREMENTAL BACKUP ........................................................................................... 21
FIGURE 10, FILE LEVEL INCREMENTAL BACKUP ............................................................................ 21
FIGURE 11, FULL BACK UP VM LEVEL ............................................................................................. 22
FIGURE 12, FULL BACK UP FILE LEVEL .......................................................................................... 22
FIGURE 13, FILE LEVEL RECOVERY ............................................................................................... 23
FIGURE 14, VM LEVEL RECOVERY ................................................................................................. 23
1 INTRODUCTION

As a focal point in every enterprise, data has a colossal measure of consideration from both scholastic and mechanical views. As data develops in size, guaranteeing that it is securely put away turns out to be difficult.

With the exponential development of Internet information services, data is a resource that ought to be kept eminently accessible. Backup is a copy of data that is required in case of a data loss or restoration. Backups protect data from human errors, hardware failures and natural disasters. The most widely recognized employments of backups are to restore the data unintentionally erased by clients and to recover from disk failures[1].

The security, reliability and scalability properties are prominent factors while recovering the data lost after a disaster to protect the critical data maintained in the enterprise. Data backup and recovery have become the essential element of data protection, as the file systems grow in size. Handling this expanding data from multiple devices like servers, desktops and handhelds effectively is of prime importance in an enterprise. The collected data may become useless when customers and organizations require this data, but is not available. This non-availability and lack of the required information may degrade the reputation of an organization. The non-availability may be the result of a disaster or human-made error. The occurrence of an unexpected event in a system leading to its disruption is a disaster.

To prevent data loss during a disaster and to back up the data, there are two approaches, which are traditional and cloud based models. Traditional model is a dedicated model while the cloud based model is a shared model. In the traditional model, the cost for the implantation is high and the recovery speed is high as well. When it comes to the cloud based model both cost and speed are lower than the traditional model[2], although traditional models have their own advantages when compared to cloud based model such as security.

An organization should have a disaster recovery plan (DRP) which is executable, testable, scalable and maintainable. This plan should achieve the target of recovery objectives with the consideration of cost constraint; that is, recovery time objective (RTO) and recovery point objective (RPO)[3]. The disaster recovery levels are characterized by these two key measures RTO and RPO which are the main objectives that need to be fulfilled when evaluating a good solution within given operating and capital costs[4].

To decrease the downtime of a system or a service after disaster, it is natural to duplicate the entire system to another standby site[5] or a recovery site away from the main regional site. However maintaining suitable standby site is quite expensive. The replicating data approach comes in three different types

i. Hot standby
ii. Cold standby
iii. Warm standby

Hot Standby is a failover mechanism, where the required data is copied to a standby site, reliably and endlessly. The standby site has to provide the required computational power, data storage and network bandwidth of previous site at all times. The major drawback of this type of method is its high operational costs.
In Cold Standby sites, the data will be copied, when there are fewer loads, such as during nights and weekends. Hence, for a situation of a disaster, the standby system must be requested, conveyed and outfitted with the last backup. The drawback in this type of approach is high RTO and RPO.

Warm standby is another replication mechanism. It stands in between hot and cold standby. The warm standby sites has an indistinguishable type of the essential frameworks like a hot standby, yet does not reflect data promptly. Rather, warm standby mechanism reproduces the data intermittently in brief time spans. Therefore, the essential and auxiliary frameworks can have small chunks of diverse information that should be reproduced in the event of a blackout. In the case of warm standby the RTO and RPO’s are less than cold standby mechanism.[2]

It is important to understand what factors may affect your Backup and DR system and how they can be improved. This thesis work attempts to provide that reliable information by conducting several experiments in an industrial setup. We have focused on performance of Backup and DR systems with disk media and cloud technologies.

1.1 Problem statement

The maintenance of a well-suited Backup and DR system in an enterprise and to have an effective approach has become a serious issue of concern[6]. Data loss can be possible due to machine failure, hardware failure and manual errors, which highly impact enterprises in prospect of budget. The aim of this thesis is to analyze the working of a current Backup and DR system with different media, investigate the problems faced by using them and recommend possible changes to their existing system by proof of concepts conducted in an enterprise. Enterprises are looking out for cloud technologies to avail advantageous offerings provided by it[7].

1.2 Objectives

The main objectives of this research study are:

1. Gain a detailed understanding of current Backup and DR system in enterprises.
2. Investigating and specifying challenges faced by enterprises with different Backup and DR systems such as traditional and cloud-based backup methods.
3. The struggles dealt with existing methodologies using different platforms such as disks and cloud in a large FMCG Company are understood by collecting data in terms of performance metrics.
4. Backup and DR will be carried out on a server with the help of a Proof of Concept, POC and test cases are run to analyze performance metrics such as RPO, RTO, Time taken for recovery, Time taken for backup, Total Cost of Ownership.
5. The experimental results are analyzed with help of data collected from enterprise.

1.3 Research Questions

1. What are the challenges faced by enterprises using different Backup and DR systems?
2. What is the time taken for backup and recovery operations in a disk-based Backup and DR system and how can they differ from a cloud-based Backup and DR system?
3. How can we assess the performance of existing Backup and DR systems in terms of performance metrics such as RPO, RTO and Total Cost of Ownership?
1.4 Outline of the thesis

The outline of this thesis is briefly described in chapter 1; introduction to this research work is explained along with its aims, objectives, research questions formulated and motivation.

In chapter 2, background of related works is explained. In background, several important terms that are required to comprehend this thesis are defined for more competence.

In chapter 3, methodology pursued throughout the work to interpret research questions is elucidated. The research strategies and experimentation are covered in this section.

In chapter 4, results obtained from survey and experiments conducted are graphically represented. The results are analyzed in this section.

In chapter 5, conclusion was defined along with how research questions were answered. The challenges that may be defeated while using a cloud-based Backup and DR are suggested.
2 RELATED WORK AND BACKGROUND

This section deals with research works that were previously accomplished. A brief description of these previous works is provided to the readers for better understanding.

Mohammad Ali Khoshkholghi et al. described that usage of cloud technologies would impact organizations on cost and speed in a beneficial manner. According to this work, crucial characteristics for any DR have been defined as minimized RTO/ RPO, privacy/ confidentiality and minimal changes in deployments at a secure site providing effective restoration rate.[2][8]

Alexander Lenk et al. indicated importance of a constructive Business Continuity Management plan to have an acceptable threshold limit of time before restoration takes place on occurrence of disaster.[9]

Aljaž et al. described that it is evident to choose an optimum Backup and DR system and it principally depends on enterprise environment. A proper decision has to be made considering requirements of enterprise such as criticality of downtime in applications, security, etc.[10][11]

Seulki Lee et al. proved from their work that proposed disaster recovery plan has high availability, continuity and low downtime with very less cost. High availability is a feature enterprises look for in any calamity. To revive and bring back business online is the point of having a disaster recovery plan essentially. Hence, high availability is highly beneficial for enterprises [13]. In most of these works mentioned, research was performed keeping in mind the needs of enterprises concerning their Backup and DR systems.

The major goal of backup is to provide data protection by creating a copy of specific record or application, which may be restored after data is lost, ruined or erased. So, backup is not an objective but it achieves protection of data. The purpose of backing up data is to empower rebuilding of information at a later point in time. Backup and archiving, are both diverse operations, which backup an auxiliary duplicate of data utilized for data protection. Conversely, archive file is the essential information, which is moved to a less-extravagant sort of media (for example, tape) in a long haul making it low budget storage [12].

Based on the backup storage media used, there are three types of Backup and DR methods majorly used; tape based, disk based and cloud based solutions. When tapes are used for backup storage purposes, it is said to be tape based Backup and DR. When disk is used as storage media then it is Disk based Backup and DR. When backup storage media used is Cloud storage, then it is said to be cloud based Backup and DR.

There are different types of backup techniques, Full back up, Incremental backup, differential backup, and the combination of these backup techniques.

2.1 Full backup

Creating a copy of an entire file system to another set of media is said to be full backup. In this type of backup, each and every folders and files are backed up on selection.

Figure 1, Full back up[13]
It is the easiest way of generating a backup. In Figure 1, all files are present in local drive and backed up to other set of media. In this type of backup, all the files backed up will be present in one place. The major advantage of this backup type is restoration of data takes lower amounts of time than other backup techniques, and it has good storage management, as backup files are stored in a single file. This being the reason, a full backup requires longer time and larger storage space. The full backup is used by companies which have less data growth from day to day[14].

2.2 Incremental backup

In this backup, the files that have been newly created or modified since previous backups are copied. It tracks and records the previous backup time, and checks time stamp while backing up and only backs up data that has been changed since the previous backup[15].

It is the fastest backup technique and requires less space as it backups only the newly created or modified data. But when it comes to recovery it requires more time and to restore individual file, initially one has to search for the file.

2.3 Full + Incremental backup

This type of backup is combination of full back up and incremental backup. First day full backup is conducted and from the next day incremental backup are performed.

![Diagram of Full & Incremental Backup](image)

Figure 2, Full and Incremental backup[13]

In this backup technique, initially a full backup is done, where each and every file is backed up. In the above figure, A, B, C, D, E, and F are different files present on a disk. During a full backup, all these files are backed up to backup media. On day 2, G
and H are added to the disk. In the incremental backup it backs up only the newly added data or modified data. Only G and H are backed up to the media. On day 3, I, J and K are added and D file is modified to the day 2 data, by doing the incremental backup D, I, J and K are added to the backup media. In this type of backup, it requires less data space in backup media and requires less time to backup. But when it comes to recovering day 3 data, all backup media from day 1 are required. Without the previous backup media, day 3 data can’t be restored making recovery process complex. This type of data can be used when there is small backup window, as it takes less time to backup.

2.4 Differential Backup

A differential backup, backs up only data that has changed or newly created since previous backup. Differential backups are cumulative, which aggregates all changes from previous backup. The data size on backup media after a differential backup grows day by day. In case of restoration, incremental backup requires each and every backup media conversely differential backup requires only the latest differential backup media for a full restoration[16].

2.5 Full + differential Backup

This type of backup is a combination of full back up and differential backup. Combinational backups reduce the space occupied on backup media and time taken for backup is diminished.

In full + differential backup, on day 1 full backup is performed, where each and every file of the file system is backed up. In above figure, A, B, C, D, E, and F are different files present on a disk. During a full backup, all these files are backed up to the
backup media. On day 2, G and H are added to disk. In differential backup it backs up only newly added data or modified data. So, only G and H files are backed up to the media. On day 3, I, J and K are added and D file is modified to day 2 data. In case of incremental backup, only D, I, J and K are backed up but in case of differential backup, newly created or modified data from last full backup is backed up, it means D, G, H, I, J and K are backed up to the media. It requires moderate data space in backup media less than full backup, slightly more than incremental backup. But when it comes to recovery, the day 3 data can be restored from backup media of day 1 and day 3. This type of backup can be used when there is less rate of change of data.

2.6 Synthetic backup

The standard backup, which is the very first copy of a full backup and the subsequent incremental or differential backups create synthetic backup. The synthetic backup is nothing but a consolidated form of existing full backup copy and consecutive incremental/differential backups[17].

2.7 File Level Storage

NAS users generally use file level storage. The file level storage is a centralized network to store all files, folders with easy access and high availability. It is comparatively inexpensive and is a storage system that enterprises look for without any complexity in management [18].

If high performance is a requirement of enterprise, then the block level storage will be more effective in comparison to file level storage.

If there is a requirement for both levels of storage then a hybrid model with both file and block level storages can be used.

2.8 Block Level Storage

Block level storage is generally used by Storage area network, SAN users. Block level storage can be compared to a hard drive connected remotely and makes use of Fiber Channel and iSCSI connectivity mechanisms of industry standards.

It is known for its pliability and adaptability as the blocks of storage volumes are created and these are individually accessed.

Block level storage has a complex manageability. It is difficult to directly access the small blocks containing files and folders[19].

2.9 VM Level Backup

In VM level backup, the complete Virtual machine is backed up. Every time there is a change, it is required to back up the whole VM and restoration from this is tedious as each file is read in the entire backup copy.

2.10 JBOD

JBOD stands for just a bunch of disks. It is a collections of disks kept at a backplane. It doesn’t have any front-end logic to manage distribution of data among disks; it represents each disk as a separate resource. JBOD can be used as direct attached storage to the host server or it can be used as storage array as in Network attached storage or with fiber channel interface can be used in storage area networks[20][21].

Depending upon the type of JBOD system using there are different types to connect disks in JBOD. Based on that there are SCSI and Fiber channel. The SATA/ SAS is
Serial attached SCSI, has a point-to-point connection with SATA/SAS port expander, which can remove the fan out problem by allowing one host controller to connect to different disks. In this thesis we have used SCSI disks.

![Diagram of SATA/SAS JBOD](image)

Figure 4, SATA/SAS JBOD[21]

2.11 AWS direct connect

AWS direct connect provides a devoted connection between enterprise data center and AWS cloud. It is different from the VPN as, VPN utilizes WAN while in AWS direct connect, a private connection is established from the on premise data center to AWS. By using this, network costs will be reduced and bandwidth throughput will be increased[22].

2.12 Performance metrics

The performance metrics we considered here are RTO, RPO, Total cost of ownership, time taken to backup and time taken to recover.

RPO and RTO are the most important recovery objectives for any enterprise prior to planning a Backup and DR solution for the company.

Total Cost of Ownership is the expenditure spent by enterprise, it plays a vital role in deciding what requirements are to be fulfilled on a undoubtedly and which may need not be of high priority

Time take to Backup and Time take to recover are compared with the same underlying infrastructure but different Backup and DR to understand which solution performed better.

2.12.1 RTO and RPO

RTO stands for recovery time objective and RPO stands for recovery point objective.
RTO and RPO are important terms related to recovery operation. RPO is the time between two backup operations. In other terms it is the maximum allowable data loss. By analyzing the RPO values we can get to know the frequency at which backup operation is performed. In asynchronous replication, data is backed up to backup media whenever there is a modification or newly creation. In this case RPO value is zero[2].

RTO is maximum permissible time for an application or data to not be available. It is the maximum time that is required for an application or a file to be available again after occurrence of a disaster[24].
3 Method

In this section, research methodology used to answer formulated research questions will be discussed giving a clear picture of experimental setup.

The purpose of research methodology is to provide scientific way to solve research questions. It includes pertinent research methods and approaches in finding solutions for research problems[25]. This master thesis has been supported by investigative approach to solve research questions and objectives. Investigative study is an important means in looking for new understanding, figuring out patterns and clearing up research problems[26]. The different research methods used are literature review, survey and case study. In the elementary stage literature review was conducted to gain knowledge of existing backup and recovery strategies and their challenges in enterprises. An online survey was conducted with professionals in the field of Backup and DR to find challenges faced by them. A questionnaire has been prepared based on literature study for an online survey. To gain better understanding of the research a case study was conducted.

![Research Methodology overview](image)

Figure 6, Research Methodology overview

3.1 Literature review

Literature review is an elementary and crucial step for research. By literature review, we can analyze, and assess problems pertinent to research area. Literature review can be based on scrutinizing scientific articles, books and other sources.[27], [28]. According to Kitchenham[28] literature review can be endeavored not only for identification of
research gap but also to epitomize existing technologies in an intensive and impartial way. By using literature review, Backup and DR technologies are analyzed and a questionnaire was prepared.

For relevant scientific articles and books various databases were referred to, which include BTH Archive Ex, ACM digital library, IEEE Xplore, Inspec/Compendex, were used. The search strategy described by Kitchenham[28] was pursued. The search without any limitations, included different whitepapers released by companies, scientific journals and online reports found on Internet.

The objective of this thesis is to find an optimal Backup and DR system for a large FMCG Company by understanding current trends followed by different enterprises. The current trends will be understood by analyzing various white papers released by enterprises. These white papers are authentic documents released by the enterprises, as current trends are scarcely present in scientific journals.

3.2 Survey

In investigative research, for accumulation of data the most common way used is survey. Information gathered from survey is valuable in creating relationship of research variables and conceivable purposes behind specific trend.[26].

The questionnaire for online survey was prepared based on literature study. It was aimed to extract different challenges faced by enterprises using their existing technologies. The interested colleagues made changes in questionnaire according to various informal discussions before conducting it online.

3.2.1 Participants of survey

For handling of survey, web based survey tool Google forms was used. It is a user-friendly survey tool. Microsoft excel and Google spreadsheets are used for the analysis of results. Answers for questions are availed in pie chart in appendix A. The participants of this survey were Backup and DR experts, working in different enterprises.

For choosing participants in the survey, we have attended several seminars conducted by Backup and DR providers.

3.3 Case study

In several procedures of research, case study in one among them. Case study is gaining comprehensive understanding of the research area, when there is a little restraint in real life scenario[29].

This thesis work was done in an FMCG Company, India. The major focus of this thesis is to find effectiveness of cloud Backup and DR for their data center. For data Backup and DR, enterprises are still using traditional methods, though Information technology world has largely adopted cloud computing. The goal of this case study is to compare traditional Backup and DR methodologies with cloud Backup and DR. For this purpose, we have used Asigra Cloud backup solution and AWS Simple storage service, S3[30] is used for data storage.

Primarily, in the case study using disk as Backup and DR medium we performed an experiment. The challenges faced while using disk Backup and DR were tried to overcome using cloud-based Backup and DR system.
3.3.1 Backup and DR using Disk

3.3.1.1 Backup server specifications

<table>
<thead>
<tr>
<th>Specification</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>Intel Xeon processor X5570</td>
</tr>
<tr>
<td>Speed</td>
<td>2.93 GHz. Turbo boost up to 3.3 GHz</td>
</tr>
<tr>
<td>RAM</td>
<td>16GB</td>
</tr>
<tr>
<td>Operating system</td>
<td>Windows server 2008 R2</td>
</tr>
</tbody>
</table>

Table 1, Server specifications of Disk Backup and DR

3.3.1.2 Test environment

The tests environment has two sites, one is primary site and other is DR site. Primary site is where enterprise infrastructure is present while DR site is established away from the primary site to relocate enterprise in case of outage at primary site. To avoid loss of data during a regional disaster occurrence or failure, there is an extra DR site located at a different location to maintain another copy of data. In above case study the DR site is maintained at a distance of 90 Kms from primary site. These two sites are connected through WAN.

In primary site, a centralized backup server is connected to another server, which is again connected to desktops, VM’s. Disk storage is connected to backup server, where data is backed up primarily. All equipment is connected in LAN. The disk storage used in this experiment in JBOD, where SCSI disks are used.

Backup server is responsible for backup and recovery operations. A backup agent is installed on server, which schedules backup operations. A backup operation is performed in specific backup window when there is less network load. Backup agent encrypts and compresses data and writes on to the disk storage. Encryption and compression techniques are chosen according to the requirement. Data backed up on disk storage is kept for 7 days, and after 7 days this data is archived on tape. These tapes are shifted to offsite for storage. At DR site, same process will be repeated as done at primary site.

Figure 7, Block diagram of Disk backup
Recovery operations are performed through backup agent, when destination is selected; data is recovered to that position. The backup agent will perform decryption and decompression. On an average, 100-120 recovery requests are made every year in the enterprise.

3.3.2 Backup and DR using cloud

3.3.2.1 Backup server specifications

<table>
<thead>
<tr>
<th>CPU</th>
<th>Intel Xeon processor X5570</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>2.93 GHz. Turbo boost up to 3.3 GHz</td>
</tr>
<tr>
<td>RAM</td>
<td>16GB</td>
</tr>
<tr>
<td>Operating system</td>
<td>Windows server 2008 R2</td>
</tr>
</tbody>
</table>

Table 2, Server specifications of cloud Backup and DR

3.3.2.2 Test environment

In this test environment, centralized backup server is connected with databases and servers. Asigra cloud solution’s backup agent[31] is used to backup data from databases and servers. This application is installed on one of the backup server. This application performs encryption and compression techniques. The backup can also be performed at a scheduled times. In this experiment we have scheduled backups at an hour level.

Backup server is connected to AWS S3 by AWS direct connect, which enables to use a dedicated connection to AWS S3. This application requires less network cost and increases bandwidth throughput[22].
4 RESULTS AND ANALYSIS

This section confers about results obtained through conducting several experiments on a disk-based Backup and DR system in a FMCG and a cloud-based Backup and DR system deployed in the same experimental setup.

4.1 Survey results

The survey results impeccably answered our research question. It introduced various techniques put to use in industries currently and their challenges faced while backing up industrial data, restoration procedures, during data loss, emergency procedures and how all operations can be done faster.

This survey gave us an opportunity to formulate a hypothesis, which was to choose the method of case study. The survey was conducted on web-based user-friendly platform, Google forms. To maximize the participation in survey, it was sent to professionals and experts who are working in Backup and DR sector. The survey was sent to approximately 300 participants who are related to cloud computing and Backup and DR Fields. Survey was open for 2 months from April 2015 to June 2015. Till the closure of survey, we have received 45 responses.

The questionnaire of survey was prepared based on literature review. And made changes by taking inputs from colleagues and experts of Backup and DR Field.

- **How frequently does your enterprise perform full back up?**

This question was asked to check whether enterprises still fear loss of data or not. By performing full backup, there is minimal risk of losing data[32]. But it needs more storage and is time consuming. For this question, we have observed that 34 respondents were performing full back up every day while 12 respondents are performing it on a weekly basis. Respondents who are using cloud Backup and DR are performing incremental-forever backup. This type of backup takes less storage and may be carried out in less time.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Everyday</td>
<td>34</td>
</tr>
<tr>
<td>Weekly</td>
<td>12</td>
</tr>
<tr>
<td>Monthly</td>
<td>9</td>
</tr>
<tr>
<td>Once every three months</td>
<td>1</td>
</tr>
<tr>
<td>No, Its incremental forever</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 3, how frequently does your enterprise performs full back up?

- **What are the major backup challenges faced by your enterprise?**

It is evident from the survey results that enterprises have multiple problems with their Backup and DR method. 30 respondents faced a challenge in backing up their remote or branch offices while 23 respondents faced a challenge to accommodate their growing data. High resource utilization is a problem for 11 respondents and 21 respondents faced a challenge of time consuming backup system. In Table 4, other challenges faced by enterprises are mentioned. To manage license agreements, which means maintaining RTO and RPO values is a challenge for 8 respondents as shown in the conducted survey.

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backup for remote offices or branch offices</td>
<td>30</td>
</tr>
<tr>
<td>Growing data</td>
<td>23</td>
</tr>
<tr>
<td>Backup is time consuming</td>
<td>21</td>
</tr>
<tr>
<td>High resource utilization</td>
<td>11</td>
</tr>
</tbody>
</table>
Table 4, what are the major backup challenges faced by your enterprise?

- **Has your company lost any critical information due to a failed recovery?**

  Out of 45 respondents, 8 respondents have lost their information minimum once. Mistakes happened while taking backup may lead to the recovery failure. Recovery failure may happen due to disk failures too.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>08</td>
</tr>
<tr>
<td>No</td>
<td>37</td>
</tr>
</tbody>
</table>

Table 5, has your company lost any critical information due to a failed recovery?

### 4.2 Experimentation results

From experimental setup 1, results obtained are graphically represented in below section. The incremental backups were performed for 7 days on both disk-based and cloud-based Backup and DR methodologies.

FMCG, which in turn is connected to one of the backup servers with configuration as shown in Figure 7 in Section 3.3.2.1. Several experiments were conducted on this machine in response to research questions.

Using the experimental setup as shown in Figure 8, values of the cloud Backup and DR are taken. As disk storage is available locally, time taken to backup and time taken to recover are lesser than cloud storage. In disk storage, the data was retained for 7 days and after these 7 days, it was archived on tapes for long time storage.

The time taken to conduct backup operation may have crucial impact on an enterprise. Size of data also appends to this factor. Incremental backups are executed for several days and each day’s incremental backup copy is repeated thrice to attain accuracy in representing results.

- **Time taken for backup**

  - **Incremental backup**

    Figure 9 and Figure 10 show the graphs of VM level and File level incremental backup. It was performed for seven days on single virtual desktops with varying data size at different intervals of time. This experiment was conducted at normal and peak hours of operations to gain better understanding of disk-based, cloud-based Backup and DR patterns. The experiment had a 3-4% of rate of change in data.
The cloud-based Backup and DR was time taking when compared to disk-based Backup and DR. Though using a dedicated channel for cloud backup, these values were obtained. Initially data is backed up to the disk. So we considered these values for disk. After 7 days the data will be archived to the tapes for storage purposes.

![VMs incremental backup](image)

Figure 9, VM incremental backup

![File level Incremental backup](image)

Figure 10, File level incremental backup

- **Full backup**

A full backup requires massive utilization values, which could not be permitted in the company leading us to run full backup operation for the same data with three iterations. Disk-based backup evidently required less time to finish the job at all three times as shown in the Figure 11 and Figure 12.
The full backup-File level was conducted on two different virtual machines installed within the same LAN network. These procedures have been performed on the same day with slight differences in time.

![Full backup-VM level](image1)

**Figure 11, Full back up VM level**

![Full backup-File level](image2)

**Figure 12, Full back up File level**

- **Time taken for recovery**
  - File level recovery

  Figure 13 and 14 show graphs of time taken for recovery operation. The recovery operation was performed to recover an accidentally deleted file from virtual machine, which was backed up 8 days back from the day of backup. For this operation, as disk
storage can store up to 7 days storage, after 7th day data was archived to tape. So to recover the file, tape level restoration had to be performed. Time taken to restore values from tapes include, selection of tapes and retrieving data from the tapes.

Time taken to recover from disk was approximately 5 times more than cloud storage.

---

![File level recovery](image1)

**Figure 13, File level recovery**

- **VM level restoration**

  To recover a crashed VM, VM level recovery operation was performed. From disk storage, the recovery operation took 100 Minutes more than the cloud storage. Time taken to locate VM snapshot on tape storage was the reason for delay to recover, as cloud storage has a management console where one can easily locate the required file.

![VM level recovery](image2)

**Figure 14, VM level recovery**

- **RTO and RPO**

  RPO and RTO define the threshold limits of an enterprise with which it can withhold the data loss and losses till business is up and running. These values were noted during the experimentation conducted at FMCG. From these values it can be understood that,
cloud backup has a better RPO and RTO values. It is difficult to restore the data of 8th day, as the local disk storage has only 7 days backup. Backed up data after 7th day is archived on to the tapes. To recover 8th day data, tape recovery has to be done. This affects RPO and RTO values. When it comes to cloud backup, it has RPO as 1hr. As backup can be done every hour. This is not possible in the disk storage, it has a constraint of scalable storage. Cloud-backup and restore solutions can be designed for multiple RPO’s and RTO’s. Disk backup and restore solutions do not have the scope to offer differential RPO’s and RTO’s. Locating the required file is also a factor of matter, which leads to increase the RPO and RTO.

Here we considered the minimum time taken for both RTO and RPO at this experimental scenario. These values may vary according to the SLA’s, Service level agreements of the enterprises where they will provide required RTO and RPO values according to their requirements.

RTO value for disk based backup and DR method has a varying value between 3 hours and 24 hours, this may be dependent on the factor of how old the data is.

Case 1: A recovery drill was performed to recover the data of Day 3 after backup. In this case the backed up data is recovered from disk. In this case RTO may be 3 hours.

Case 2: Recovery drill is performed to recover the data of Day 8; in this case the data has to be recovered from tapes. So the time taken to recover will be more. In our experiment RTO in this case is 7 hours.

Case 3: A recovery drill was performed when a disaster occurred at primary site. In this case data has to be recovered from the DR site. In this case the RTO will be up to 24 hours.

<table>
<thead>
<tr>
<th>Type of Backup</th>
<th>RTO</th>
<th>RPO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disk based</td>
<td>3 to 24 hrs.</td>
<td>4 hr.</td>
</tr>
<tr>
<td>Cloud based</td>
<td>2 hrs.</td>
<td>1 hr.</td>
</tr>
</tbody>
</table>

Table 6, RTO and RPO values

- **Total cost of ownership**

  The total cost of ownership for disk and cloud were taken by considering the infrastructure cost that needs to be set up for both types of backup. As cloud backup doesn’t require any high-end capital expenditure. It only focuses mainly on operational expenditure. The operational expenditure for cloud backup is given in table 7. While the disk backup needs capital expenditure that needs to be setup before performing the operation. Along with the capital expenditure, there is operational expenditure involved too. By considering all this, cost for disk based Backup and DR was around $6 more than cloud based Backup and DR.

  For disk backup and DR method, Capital expenditure and operational expenditure were considered. Both Capital and operational expenditure were given by the enterprise. In cloud backup and DR method, only operational expenditure was considered because there was minute capital expenditure to be invested.

<table>
<thead>
<tr>
<th>Type of backup and DR</th>
<th>Price (Per GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disk</td>
<td>$7</td>
</tr>
<tr>
<td>Cloud</td>
<td>$1.2</td>
</tr>
</tbody>
</table>

Table 7, Costs per GB
4.3 Analysis

- **Time taken for backup**
  Time taken for backup was more in cloud based Backup and DR than disk based backup, as the disk backup is located in a local network. So the time taken for backup was less.

  Though, in our thesis we have used AWS direct connect time taken was more. Storage region was selected such a way that it is nearest to the backing up region.

- **Time taken for recovery**
  Time taken for recovery was more for disk based backup, though the disk storage was locally present. The major problem faced during recovery from disk was to locate file in disks. The disk storage can store data up to 7 days from the day of backup. After 7 days, data was archived into tapes to cope up with the growing data.

  To recover data after 7 days, time taken will be more, as recovery operation has to be performed from the tape. This is in case of accidental deletion of data of hardware failure.

  But in case of cloud, time taken was less. Cloud storage has a management console where one can easily select the required data for restoration.

- **RPO and RTO**
  These values are less for cloud backup, as locating data in cloud storage was easy. When there is an outage at primary site, data has to be recovered from the DR site. In cloud storage it works as both primary storage and DR storage. But in case of disk, data has to be recovered from the DR site, which is 90 KMs away from primary site. This affects time taken for recovery.
5 CONCLUSION

In this thesis, we firstly try to understand current technologies prevailing in industry sector through literature study. To compliment this knowledge and gain more perspective from a real time scenario, a questionnaire was prepared which renowned experts of Backup and DR system from all over the world answered.

A case study with disk-based Backup and DR was further conducted and analyzed on its performance scale while comparing these values with a cloud-based backup and disaster recovery system which was deployed on the same network.

This research dealt with understanding better methodology for Backup and DR in an enterprise. In beginning, personal interviews were conducted with professionals in Backup and DR industry for decades.

Backup and DR has been the way to store data for making a copy, protecting it in a secure site to avoid loss of data in case of disruptions with the normal course of actions in enterprises. Beginning of backup technology was by storing hard copies in warehouses. Later, with invention of magnetic media, backups were slowly stored in tapes for many years. To avoid closure of companies during occurrence of disasters, initiation for saving prior data was made into which efforts were put in.

In this thesis we found that time taken for backup was more in cloud backup than disk backup, when it comes to time taken for recovery. RTO, RPO and total cost of ownership was more in disk backup. When an enterprise considers RTO and RTO as their prime factors to choose a backup and DR method, then the enterprise may opt for cloud backup and DR method.

Parameters such as recovery point objective, recovery time objective, time taken for backup and time taken for recovery have been compared with a cloud-based Backup and DR system, which has been deployed in the FMCG Company.

By the case study performed on disk backup, various challenges were faced; these challenges were overcome by using cloud Backup and DR.

- **Growing data**

  There is scalability issue with the disk storage. When there is an unexpected growth size of data in the one-day backup, disk needs to be placed accordingly. Cloud storage can handle unexpected growths.

- **Remote locations and branch office backup**

  Traditional backup works well in companies where the data is mostly concentrated in one location or data center. As companies get more global, their backups get siloed into point solutions, making it difficult to manage. As data gets dispersed in many locations in global organizations, cloud-based backup makes it easier to consolidate and manage all backups through a single pane of glass.

- **Recovery failure**

  Cloud-based backup solutions have better autonomic healing features that ensure that backups are actually restorable. Disk-based traditional methods usually do not offer features that guarantee recoverability of backed up data, rendering backups potentially useless. There are balanced chances of data loss due to disk failures and there won’t be any loss of data due to recovery failure.

- **Disk backup is a complex process.**

  Many people are involved, making it error-prone. In small branch offices, office administrators usually do backup, not IT personnel, making backups unreliable. While Cloud-based backup is more automated and less manpower-intensive, making it less expensive and less error-prone.

- **Cloud-based solutions have features like multi-tenancy, encryption and compression built in.**

  Disk backup and recovery solutions do not offer multi-tenancy, encryption and similar features.
There are several failures that have occurred in cloud storage. In June 2010, Hosting.com data center was down for few hours, as there was software vulnerability. In April, 2011 Amazon Elastic Block Storage, EBS and RDS services were down for 4 days[14]. To have a better availability there is a requirement to backup to different regions.

- **Future work**

  This experiment was performed using Asigra cloud backup solution, it can be further evaluated using AWS storage gateway. Gateway stored volumes and Gateway cached volumes can be used which will reduce the latency by storing the frequently accessed data locally. There is a possibility to use a hybrid model, where cloud storage can be used with either tape or disk storage media locally.

- **Answering the research questions**

  - **What are the challenges faced by enterprises using different Backup and DR systems?**

    This is answered by conducting survey. Experts of Backup and DR and employees of different companies related to backup have answered this survey, results of which are available in Appendix A. According to the survey, major backup challenges were to take backup for remote locations. This challenge was even faced when performing case study. But cloud Backup and DR overcame this challenge.

  - **What is the time taken for backup and recovery operations in a disk-based Backup and DR system and how can they differ from a cloud-based Backup and DR system?**

    The time taken to conduct a full back up in two environments with disk-based, cloud-based Backup and DR environments set up in an experiment conducted at a FMCG Company was attained. Disk-based Backup and DR system was proved to be better with slight variation in the time taken to complete backup operation.

    Incremental backups were conducted in two methods. Virtual machine incremental backups and full incremental backups were conducted as shown in section 4.2

    In the case of virtual machines incremental backup, disk-based backup was exceptionally faster when compared to that of cloud-based backup methodology. But, in full incremental backup, the changes in time were very minute.

    In case of recovery, time taken for recovery was more for disk backup method than the cloud backup and DR. The reasons were discussed in section 4.2 and 4.3

  - **How can we assess the performance of existing Backup and DR systems in terms of performance metrics such as RPO, RTO and Total Cost of Ownership?**

    The RTO and RPO values were retrieved. Cloud based Backup and DR had better RPO and RTO values than disk Backup and DR. As there are several challenges and time taking tasks that need to be performed for recovering data from disk. The Total Cost of Ownership was higher in disk-based when compared to the cloud-based Backup. Various parameters contributing to this costs in a Backup and DR plan are mentioned in Appendix B
REFERENCES


[15] A. the A. J. H. J. H. is an A. I. S. in G. S. D. J. joined I. D. in 2005 as a S. A. with focus on T. S. M. F. some years he was responsible for the customers in the shipping industry M. recently and he has been working as T. L. in B. S. I. the delivery centers in Denmark, “Incremental forever backup versus traditional backup,” *IT Service Management 360*. .


Cloud Backup: Advantages and Disadvantages Compared with Traditional Methods. | Esys - Technology, Medical and Research.


“JBOD Vs RAID : Cost Effective Network Storage Options » JetStor.”


APPENDIX A

The survey conducted had the following responses as shown below.

What method is used to backup data in your company?

- Disk-based: 42%
- Tape-based: 2%
- Cloud-based: 16%
- Tape/Disk: 5%
- Tape and Disk: 2%
- Disk and Cloud: 11%
- Virtual Disk: 2%
How frequently do you perform a full backup of data?

- No, it is incremental forever
- Once every three months
- Monthly
- Weekly
- Everyday

The primary problem (if any) of the data backup and disaster recovery method used by your company is?

- No problems involved, satisfactory
- Not able to accommodate user data
- Restoration automation
- Complex to scale and manage
- Managed by third party
- Not reliable
- Not secure
- Expensive
- Not efficient
- Not fast enough
Has your company lost any critical information due to a failed recovery?

The data backup is always carried out within the expected time frame (within the backup window)?
What are the major backup challenges faced by you?

- High resource utilization
- Data backup requires more storage space
- Growing data
- Management of license agreements
- Total cost of ownership
- Backed up data security
- Backup is time consuming
- Backup for remote offices or branch offices

How do you decide which files are to be destroyed and which ones to be saved?

- Based on retention and destruction policy
- We informally delete data based on its requirement
- Deletion of data when storage is full
How frequently do you recover the data from the backup?

- Only during failures
- Once in a year
- Once in six months
- Once every three months
- Monthly
- Weekly
- Everyday

According to you, do you think your backup and recovery process is very expensive?

- Yes
- No
Does your company perform disaster recovery (DR) drills?

Are you completely satisfied with the security provided by your data backup and disaster recovery system?
Does your company use compression, encryption techniques while backing up data and does it support multi-tenancy?

The computing resources required for the data backup and disaster recovery is?
APPENDIX B

These factors contribute to Total Cost of Ownership in an enterprise. Depending on requirements of an enterprise, following factors may be included or excluded from their Backup and DR plan.

Total Cost of Ownership- Constituents of different costs for Tape/Disk Backup

- Capital cost
  - License cost + base/server cost
    - Hardware purchase
    - Software purchase
    - Local and remote data circuits
    - Storage area networking
    - Security and Encryption
    - Compression
    - DR site costs
    - Client side cost
    - Cost of growth
  - Integration
    - Backup infrastructure
    - Backup media
    - CIFS (Common Internet File System) or NFS (Network File System)
  - Migration
    - Migration (lifecycle costs of the storage system), remastering (data lifecycle costs)
- Operational cost
  - Training
    - Backup and disaster recovery labor (DR planning and testing)
  - Insurance
  - IT staff
    - Storage management labor (upgrades, troubleshooting, load balancing, tuning)
    - Monitoring costs
  - Management time
    - Cost of disaster risk, business resumption
    - Recovery time objective and recovery point objective (RTO and RPO)
  - Electricity
    - Power consumption
    - Cooling
  - Floor space
    - Data center floor space
  - Outage costs
    - Cost of scheduled outage
    - Cost of unscheduled outage (machine related)
    - Cost of unscheduled outage (people and process related)
  - Backup and recovery cost
    - Hardware maintenance
    - Software maintenance
    - Cost of performance
  - Cost of procurement
  - Transportation costs
- Risk cost
- Cost of waste
- Cost of duplicate data
- Data loss - Loss of reputation (which is immeasurable)
- Litigation, e-discovery risk (lawsuits)
- Reduction of hazardous waste
- Cost of risk with backup windows
- Noncompliance risk (archive, data retention) - negative publicity

- **Opportunity cost** (opportunity cost is the value of the opportunity lost).