# Survey on Data Storage and Backup Methodologies

**[1]Tejashree Ghadge, [2]Dr. Swarnalatha P**

[1]M. Tech Computer Science and Engineering, [2]Associate Professor Computer Science and Engineering
School of Computer Science and Engineering,
VIT University, Vellore, India

*Abstract*— **In today's world more than five billion Gigabytes of data is generated every ten minutes. Managing and protection of the tremendous data becomes major concern as data can be lost because of hardware failure, man made mistakes or unpredictable nature. Protection of data requires taking backup and in order to take backup various traditional technologies such as disks, tapes, optical media, NAS, DAS and virtualization is used. Major drawbacks of traditional technologies are hardware failures, requirement of more storage space and security. Hence new technologies such as storage virtualization provides efficient in storage space, less time for backing up data and more security. This paper explains the survey of different technologies used for backing up data on storage.**

*Index Terms*— **Disks, tapes, NAS, DAS, storage virtualization.**
_____

## I. INTRODUCTION

In today's world due to exponential growth of data due to social media, online accounts, data is growing at the faster rate. The amount of data produced in 2003 was around 5 gigabytes. The same amount of data is being produced within two minutes now-a-days. Hence there is need for managing and protection of these data. In order to protect data backups are performed in case of disaster or some failure. Tapes and disks were used to previously to perform backups. The backups executed were lengthy and needed more space that means they were time and space consuming. As the data is growing tremendously, there was need to move to faster and time saving option. Also tapes and disks required more cost and are less secure. Hence there was a need to go for virtualization storage backups. Most of the organizations these days use virtual storages. Also the backups done are faster and require less memory. Snapshots of the memory are taken in order to backup data. These are not actual backups but store the memory location of the actual data. Therefore the time require taking snapshot or a snap is less and also the space required to store the snapshots are less as they store only the pointers of the memory location. Hence more advanced technology is used today for performing backup of the data. This will help the user to store more data and time required to store and restore his data will also be less.

## II. TRADITIONAL STORAGE SYSTEMS

### A. Tape Backup

In tape drives data is stored on the magnetic tapes where read and write operations are performed. Tape backup is backing up the data from the storage device to the tape in order to recover it from the hardware failure. Tape backups can be done manually or automatically with the help of software. Tape backups are expensive and with less storage space. Tape backups are usually used for offline backup of data where user cannot access the data while the backup operation is performing. Tape drives allow sequential access storage and hence backup requires more time. The image displayed shows how tapes are used for storing and backing up different data files.



**Fig.1. Tape Backup**

Advantages of the tape drives are; ease of use as tape drives have lots of support. Multiple redundant backup set are used in tape backups. Tape backups are reliable as tapes are used only during backup. Ease of storage as tape cartridges are small and can be

easily stored offsite. Disadvantages of tape drives are; more expensive as tape drives are more costly than hard disks or network backup. Tape degradation as magnetic tapes gets easily affected due to heat, dust, mishandling. Tape backups are less secure than disk drives.

### B.  Disks Backup

In disk storage data is stored in the form of magnetic, electronic and mechanical on the surface of rotating disks. When large amount of data is stored on the disks storage systems it is known as disks backups. Backing up data to disks is used for data replication and disaster recovery. Disks backups are faster than tape based backups. Traditional backups were carried out using tape media but the drawback of tape backups was that they were slower compared to disk backups. Also restoration of data was unreliable, because of the delicacy of tape and also because searching of data was difficult unless the user knows the tape and file names required. Loss of data and security are also problems commonly associated with tape.



**Fig.2. Disk Backup**

Advantages of Disk backup are; faster than tape backups as disks backups allows administrator to perform backups more efficiently and quickly than from tape. Disk backup provides data de-duplication. Disk system offers either data de-duplication at the source side or at the destination side. The de-duplication on the destination side is faster and it requires less overhead on the source. Disk backup provides data compression. It allows storage of more data on less disk space. Disk backups are more reliable. This is highest advantage over tape disk failures and its inability to restore data.
Disadvantages of disk backup are; the major concern for disk backup is more cost for storage and backing up data compared to new technologies. It is not archived storage as it doesn't provide backups with more permanent storage systems. When disk fails whole of the data is lost hence it does not provide with double disk failure.

## III. DAS, NAS AND SAN STORAGE BACKUP

### A.  Direct Attached Storage

DAS is Direct Attached Storage where application server is directly attached to a disk. The common example of a DAS is internal hard disk in a laptop or a personal computer. It can be explained as single or multiple arrays combined together as in RAID (Redundant Array of Independent Disks). In DAS user is connected to the server with the help of host bus adapter. DAS is used to backup data at very small scale. The figure shows that applications are attached directly to data storage.
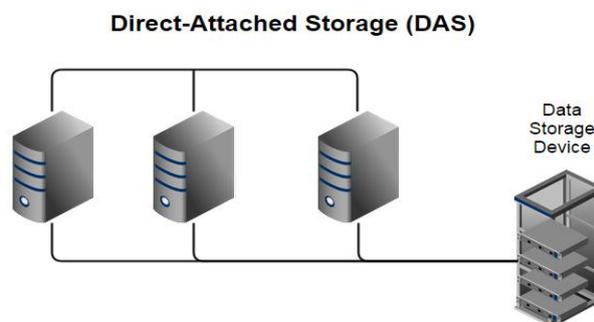


**Fig.3. Direct Attached Storage**

Advantages of DAS are; DAS is simple and cost efficient. DAS is used to run small applications like financial accounts, emails and database applications on server. DAS provides both block level access and file level access for end users.

Disadvantages of DAS are; Degradation in performance when trying to access multiple applications at the same time. DAS provides limited scalability. It is not helpful for the applications which require immediate scaling of data. As end user/client has to be reliable on its server ad if the server is down or it has any viruses user will be able to access the data.

### B.   Network Attached Storage

NAS is network attached storage where applications are attached to file systems or servers with network as interface. Usually Ethernet switches are used to attach applications to the server. NAS is file level access means file are used at elementary level for read write of data or to take backups of the data. NAS is used for small and medium sized organizations so that one gigabit of ports is connected to provide high performance. The figure explains that applications or clients are attached to servers using Ethernet switch.
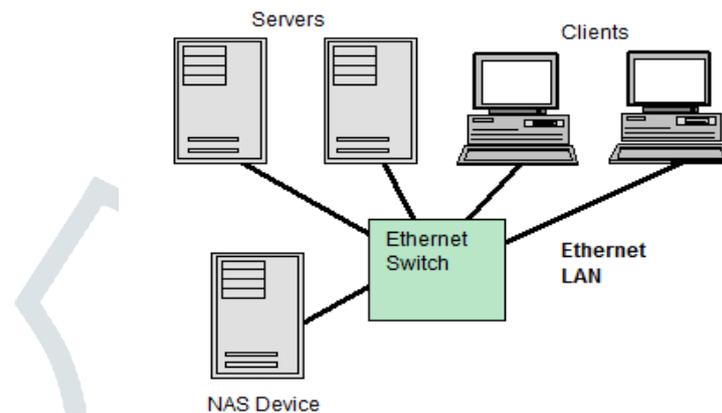


**Fig.4. Network Attached Storage [1]**

Advantages of NAS are; NAS is simple. It provides guided wizards for setup and to quickly add storage to network. NAS uses Linux as operating system hence reduces the licensing cost provided for windows. NAS provides RAID support. RAID is used for storing data on multiple disks hence it is more fault tolerant and used for disaster recovery. Data can be backed up on multiple disks using RAID which is useful in single disk failure.

### C.   Storage Attached Network

SAN is Storage Attached Network where multiple applications are connected over a network to multiple storage devices. San is block level access as the read write operations are performed at the block level. They are mostly used for backup and disaster recovery mechanism as they multiple servers are used. In SAN data can be transferred from one storage system to another storage system without servers and hence reduces the CPU cycle time for backups and are faster. SAN uses FC (Fiber Channel) technology in order to transfer data over large geographically located servers. It allows IT organizations to backup their data at remote locations. The figure shows the application servers are attached to RAID array using fiber channel.
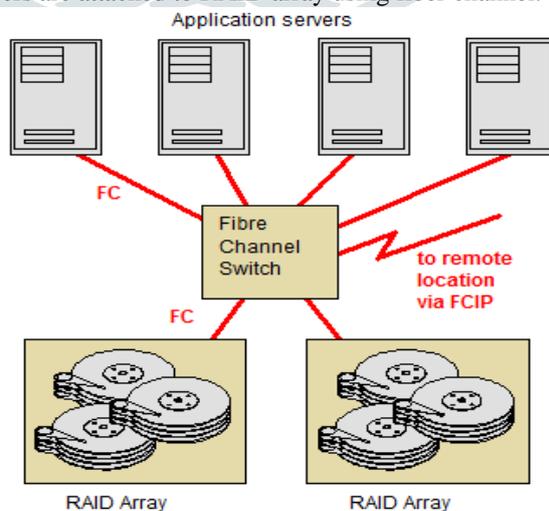


**Fig.5. Storage Attached Network**

Advantages of SAN are; SAN provides block level access hence backing up of data requires less time. As SAN removes storage from servers more storage utilization is achieved. Hence there is no need for additional storage hardware and therefore saves

money. Disadvantages of SAN are; backup consumes more memory, I/O bandwidth etc. In SAN there can be issues for restoration from remote servers. More cost required for Fiber channel transfers.

## IV. VIRTUALIZED STORAGE BACKUPS

There are many drawbacks in disk based backups as increase in data continues to outperform the backup infrastructure and is unable to accommodate new requirements. This has led to difficulty in meeting backup windows and a lack of transparency about the efficiency of backups. A new technology such as snapshot technology has emerged to fast and time efficient backups. Most the enterprises now use snapshot technology to backup their critical data.

### A. Snapshot for backup

A snapshot is a point in time copy of the data. As the snapshot file is only a change log of the original virtual disk, snapshots are not actual backups. Snapshots are time and space efficient. It stores the pointer to the memory/storage of which backups are to be taken. If the files are modified inside the volume the changes are stored in the snapshot with new pointer to the changed data. Many organizations such as VMWare, NetApp, Symantec, IBM, and EMC[2] use snapshot technologies but not all have equality. Amongst the above NetApp is the first company which allows volume level snapshot which takes point in time copy of the volume. Below figure gives the detailed view how snapshot technology works in NetApp filers [1].
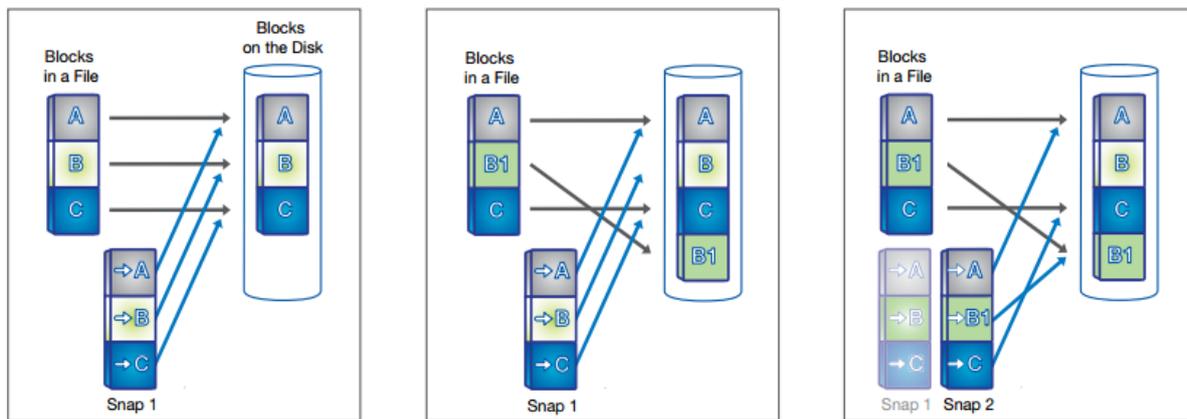


| Figure 1 a | Figure 1 b | Figure 1 c  [2] |

Above figure explains that A, B and C blocks are stored on NetApp storage as backups. When snapshot is taken pointers to the memory locations are stored in Snap 1. When file B is modified the new changed copy of the file is stored as B1 and Snap 1 stills points to old blocks of data. When second snapshot is taken Snap 2 contains the pointers to the changed blocks as shown in Figure 1 c. Hence only the modified block requires new memory and not all the blocks are copied. This allows saving storage space and time required to take the snapshots also reduces. Also the restoration of data is easier as we have to refer only to the snaps which are to be restored.

## V. COMPARISON BETWEEN TRADITIONAL BACKUP; DAS, NAS, SAN BACKUP AND VIRTUALIZED BACKUP

### A. Time and Space

The time required for performing backup plays major role as now-a-days all users look for faster technologies. Newer the technologies used to perform backups, lesser the time required. Tapes are the slowest for backing up the data. Also tape backups require more space and memory for storing small amount of data. Disk backups are faster compared to tape but slower than NAS and virtualized backups.  NAS/DAS/SAN are used for small organizations or over the network where much storage is attached over the switch or Ethernet. They are relatively quicker than tape and disks backups. The fastest backup is using virtualized backup technology. It does not actually require the storage space to backup but only needs the memory space where changes are made. It uses snapshot technology to backup data. It is the fastest technology used to perform backups. It only stores the pointers to the memory location and the pointers are changed after taking next consecutive snapshots. The only memory required is to store the changes made in the original file. Hence they are most time and space efficient.

### B.    Reliability and cost

As user's data is sensitive it must be stored on trustworthy and reliable systems. Also cost must be less and data should be secure. Tapes are more prone to environmental conditions and easily degradable; hence are least reliable. Disks are costly and are not as much of reliable. When one disk fails and entire data is stored on that disk, complete data will be lost. Hence it requires double disk failures. In DAS/NAS/SAN RAID disks are used, which provides more reliability for data. There are issues for restoration of data and also the fiber channel cost required is more. Snapshot technology requires less space and also less time. They are more reliable than all the backups.

## VI. CONCLUSION

In today's world data is increasing at the exponential speed and hence protection of data is major concern. There are many traditional ways in which data can be protected but most of them have become obsolete and are not useful in large organizations due to new emerging technologies. There are numerous disadvantages in using traditional storage systems for backup as speed required and space required is more and hence there is need to switch towards new technologies like virtualization. Now-a-days most of the companies use snapshot technology for backing up their data. Snaps stores only the pointers of the memory location and hence are faster than older methodology of backup. This shows that in today's world virtualization plays major role in data protection and management.

## REFERENCES

[1]  http://encyclopedia2.thefreedictionary.com/Network-attached+storage
[2]  http://www.ant.hu/sites/default/files/u3/netapp_snapshot_ds-2477.pdf
[3]  A.FARITHA BANU, C. CHANDRASEKAR, "A SURVEY ON DEDUPLICATION METHODS", International Journal of Computer Trends and Technology- volume3Issue3- 2012
[4]  Mark W. Storer Kevin Greenan Darrell D. E. Long Ethan L. Miller," Secure Data Deduplication"
[5]  Whitepaper on EU Data Protection, "Amazon Web Services EU Data Protection Whitepaper"
[6]  Michelle M. Howell. "Data Backups and Disaster Recovery Planning"
[7]  https://www.backupassist.com/education/articles/tape-backup-introduction.html
[8]  IBM Corporation, Hard Disk Drives (HDD) Web Page, http://www.pc.ibm.com/ww/eserver/xseries/hdd.html, accessed November 3, 2003