ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 2, March 2013

# A Survey: Data Storage Technologies

Pradnya Eknath Gaonkar, Sachin Bojewar, Jayesh Ajit Das
PG Student, VIT, Mumbai, Maharashtra, India
Assistant Professor, Computer Engineering, VIT, Mumbai, Maharashtra, India
BI Consultant, Maia Intelligence, Mahape, Navi Mumbai, Maharashtra India

***Abstract*— nowadays data is increasing rapidly, so to store these data we need data storage technologies .In this survey paper we have calculated estimated growth over 10 years of increasing volume of data. Also have compared data storage technologies like Direct Attached Storage (DAS), Network attached Storage (NAS) and Storage area Network (SAN).We have focused on emerging data storage technology such as cloud storage and included some strategy for removing unwanted data. There is inclusion of advance data storage devices such as holographic data storage and memory stick.***

***Index Terms*—Cloud Storage, Direct Attached Storage (DAS), Network Attached Storage (NAS), Storage Area Network (SAN)**

## I. INTRODUCTION

The world's information is doubling every two years [1]. In 2011 the world will create a staggering 1.8 zettabytes. By 2020 the world will generate 50 times the amount of information. We always knew it was big in 2010 cracking the zettabyte barrier. In 2011, the amount of information created and replicated will surpass 1.8 zettabytes (1.8 trillion gigabytes) - growing by a factor of 9 in just five years. While 75% of the information in the digital universe is generated by individuals, enterprises have some liability for 80% of information in the digital universe at some point in its digital life. The digital universe is something to behold 1.8 trillion gigabytes in 500 quadrillion "files" and more than doubling every two years. The amount of information individuals creates themselves — writing documents, taking pictures, downloading music, etc [2].Data here means a lot. Not necessarily a text message or a spreadsheet. Audio, Video included. If data is divided into two different types, as structured data and un-structured data, the growth of un-structured data is posing a huge challenge. It is reported that the world is crossing the zettabyte barrier meaning that in the year 2010, the total digital data is exceeding 1 zettabyte [3]. Table I shows data unit.

**Table I Data Unit**

| | |
|---|---|
| Binary digit | 1 bit |
| 1 Byte | 8 Bits |
| 1 Kilobyte | 1024 Bytes |
| 1 Megabyte | 1024 Kilobytes |
| 1 Gigabyte | 1024 Megabytes |
| 1 Terabyte | 1024 Gigabytes |
| 1 Petabyte | 1024 Terabytes |
| 1 Exabyte | 1024 Petabytes |
| 1 Zettabyte | 1024 Exabytes |
| 1 Yottabyte | 1024 Zettabytes |
| 1 Brontobyte | 1024 Yottabytes |
| 1 Geopbyte | 1024 Brontobytes |
| 1 Epic Byte | 1024 Geopbytes |

## II. MOTIVATION

Information storage is a central pillar of information technology. A large quantity of digital information is being created every moment by individual and corporate consumers of IT. This information needs to be stored, protected, optimized, and managed [4].Every day, we create 2.5 quintillion bytes of data - so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 2, March 2013

climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few [5].

## III. DATA DISTRIBUTATION STSTISTICS

### A. *Distribution of 1 GB data*

Table II shows some basic consideration of medium scale company which generates below mentioned volume of data and Fig. 1 shows the pie-chart for distribution of 1GB data.

**Table II Distribution of 1GB data**

| Sources | Data in GB |
|---|---|
| Development | 0.2 |
| Database | 0.05 |
| CCTV | 0.7 |
| Records | 0.05 |
| Total | 1 |

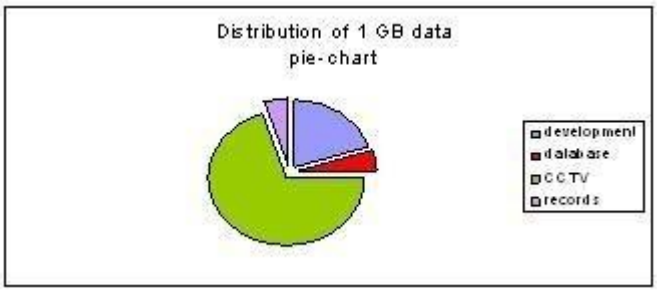


**Fig. 1 Pie-chart for distribution of 1GB data**

### B. *Formula to Calculate Data Growth*

Formula: (The world's information doubling every two years [1].)

$$y = a(1+r)^x \qquad (1)$$

Equation 1 [6] is use for calculating data growth where y, a, r and x are refer as data growth, initial value of data, data rate and required year respective. To calculate value of data rate (r) assume values of data growth (y) =2 GB, initial value of data (a) =1GB and required year (x) =2 years. After substituting values of y, a and x we get value of data rate (r). So r=0.4142% from initial value (a) and data rate (r) we can calculate data growth (y) after required year(x).

### C. *Increasing Volume of Data Statistics*

Table III shows the data growth over 10 years. We have data rate (r) = 0.4142%.After substitution of initial value (a), data rate(r) and required year (x) which varies between one to ten year .We can calculate data growth over 10 years and Fig. 2 shows exponential data growth after 10 years .

## IV. DATA STORAGE TECHNOLOGIES

### A. *Direct Attached Storage (DAS)*

In Direct Attached Storage, Storage is directly attached by a cable to the computer processor. I/O requests access devices directly. Storage system directly attached to server or workstation, without a storage network in between. There are various advantages of DAS. First, fast file transfers. With up to 320MB/sec transfer rate, DAS is currently the fastest storage solution available today. Second, high security.DAS is also very secure since it utilizes a direct physical connection to the server, and cannot be spoofed (tricked) into giving access to unauthorized sources. Some disadvantages of DAS, first, limited scalability. A host bus adaptor can only support a limited number of drives. For environment with stringent up-time requirement, or for environment with rapidly increasing storage requirement, DAS may not be the right choice. Second distance limitation. SCSI device connections cannot typically exceed 12 meters, which means data storage is usually confined to a single room, or even within a single system enclosure [7]. Fig. 3 shows the direct Attached Storage.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 2, March 2013

**Table III Calculation of 1 GB Data growth after 10 year**

| Years | Data | Initial Data ( a)=1 GB |
|---|---|---|
| 1 | 1.4142 | 1 |
| 2 | 1.99996164 | 1 |
| 3 | 2.828345751 | 1 |
| 4 | 3.999846561 | 1 |
| 5 | 5.656583007 | 1 |
| 6 | 7.999539689 | 1 |
| 7 | 11.31294903 | 1 |
| 8 | 15.99877252 | 1 |
| 9 | 22.62546409 | 1 |
| 10 | 31.99693132 | 1 |

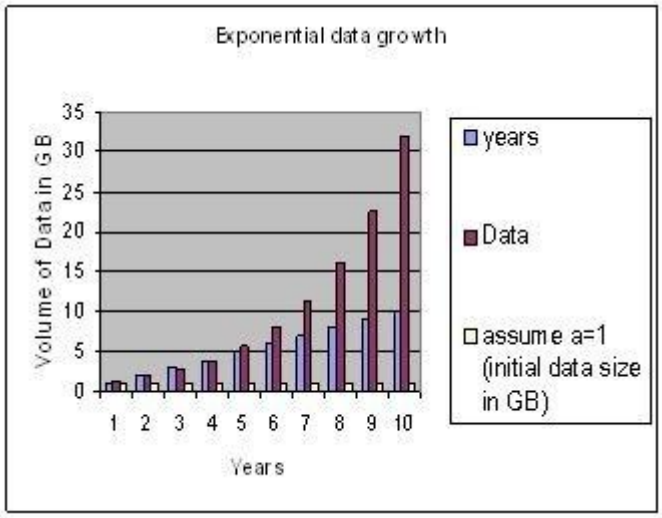


**Fig. 2 Bar chart for exponential data growth after 10 years**

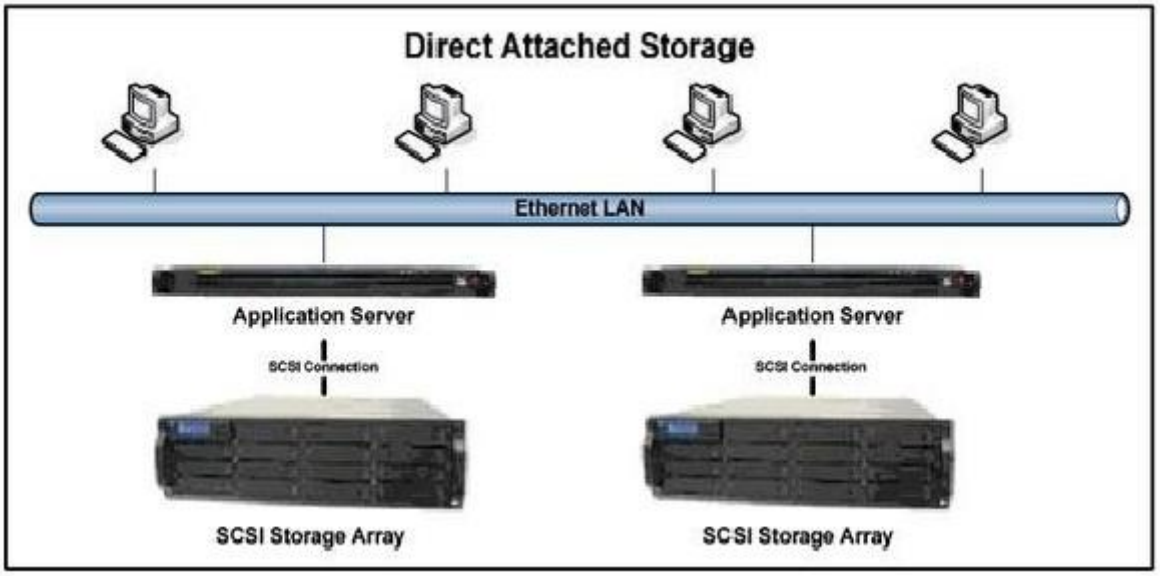


**Fig. 3 Direct Attached Storage [7]**

### B. *Network Attached Storage (NAS)*

Network attached storage has become a popular storage solution. NAS appliances are storage devices that connect directly to the LAN via standard Ethernet port and use the familiar TCP/IP protocol to communicate with network peers. TCP/IP works by dividing up actual files into many small fragments, encapsulating into packets, and then sending as frames through the LAN or WAN (Wide Area Network).Advantage of NAS over other storage technologies is its ease of implementation.The plug and play nature of NAS makes it a flexible storage option for small businesses and is easily administered by existing IT staff. In addition, the use of TCP/IP eliminates distance limitations associated with DAS storage solutions [7].Multiple users can access the storage drive at the same time so files can be shared among multiple users and devices. Among the disadvantages of NAS file transfer rate is not as fast as DAS.NAS always suffers from network overhead, NAS device is often used by many computers on a network, it can be overwhelmed by everyone trying to access the device at the same time. This could be caused by limited network bandwidth, and the processing power within the device. Another disadvantage is that the NAS device could become single point of failure, especially if it becomes a central data storage unit within an organization. Solutions to this issue would be to distribute important files across multiple units, and to schedule and run regular remote backup routines on these machines [8]. Fig .4 shows Network Attached Storage.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 2, March 2013

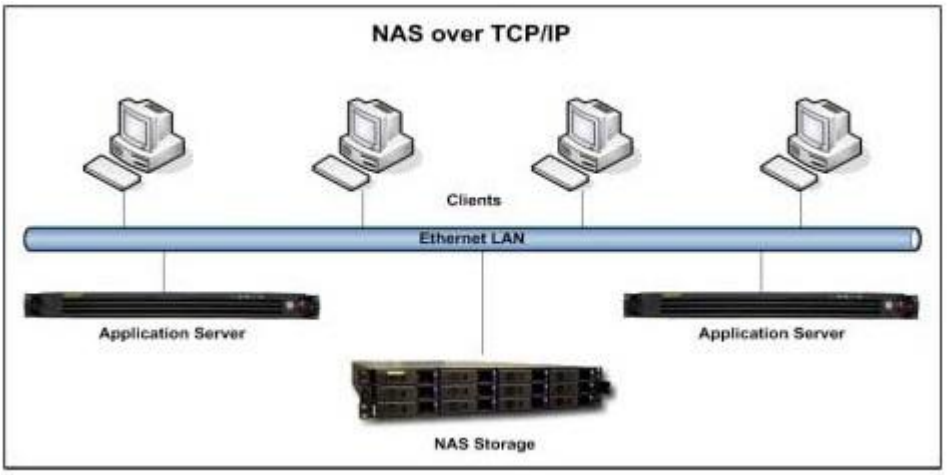


**Fig. 4: Network attached storage [7]**

### C. *Storage Attached Network (SAN)*

Storage Area Network is a network dedicated to providing storage to enterprise servers. SANs are traditionally configured using special switches and storage devices that communicate via Fibre Channel protocol. The Fibre Channel protocol uses block-level SCSI commands which are transmitted over serial, rather than parallel connections, and can span up to 500 meters. Servers connect to the SAN using special Fibre Channel HBAs (Host Bus Adapters) capable of transmitting data at 1Gbps, 2Gpbs, and soon even 10Gbps. SANs also offer a centrally accessible/manageable data store easing administration [7]**.**Advantages of SAN are first, it provides high scalability any number of storage devices can be added to store hundreds of terabytes. Second it's have ability to add additional capacity. The primary disadvantage of SANs is cost. Since it is still a rather new technology, Fibre Channel equipment remains expensive, as is the support required to implement and maintain it [7].

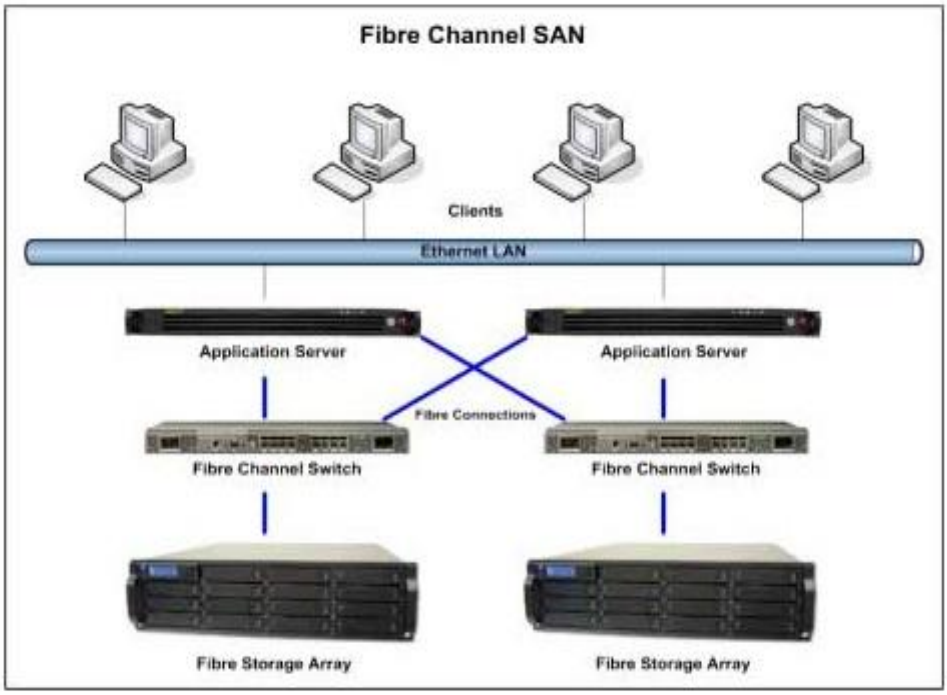


**Fig. 4 Storage Area Network [7]**

## V. COMPARISION BETWEEN DAS, NAS AND SAN

Table IV shows the comparison of storage techniques across various parameters.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 2, March 2013

**Table IV Comparison between DAS, NAS and SAN**

| Storage Characteristics | DAS | NAS | SAN |
|---|---|---|---|
| Centralize Management | No | Yes | No |
| Storage Focus | Server Centric | Network Centric | Channel Centric |
| Network Used | No | Yes | Yes |
| Data Sharing | No | Yes | Yes But requires specialize software |
| Interface | Directly attached to server | Attached on the primary LAN | Acts as a secondary network to LAN |
| Appearance to user | Storage is directly attached to computer via cable | NAS appears as a single node on network | SAN appears as extra storage for each server |
| Storage scalability | No | Yes | Yes |
| Storage units separate from servers | No | Yes | Yes |
| Network bottlenecks avoided | Yes | No | No |
| Single server bottlenecks avoided | No | Yes | Yes |

## VI. EMERGING TECHNOLOGY FOR DATA STORAGE

### A. *Cloud Storage*

Cloud storage is a model of networked online storage where data is stored in virtualized pools of storage which are generally hosted by third parties. Hosting companies operate large data centers, and people who require their data to be hosted buy or lease storage capacity from them. The data center operators, in the background, virtualized the resources according to the requirements of the customer and expose them as storage pools, which the customers can themselves use to store files or data objects. Physically, the resource may span across multiple servers [9]. Cloud data storage has many advantages. It's cheap, doesn't require installation, doesn't need replacing, has backup and recovery systems, has no physical presence, requires no environmental conditions, requires no personnel and doesn't require energy for power or cooling. Cloud data storage however has several major drawbacks, including performance, availability, incompatible interfaces and lack of standards [10].Another advantage of using online services to store your files is ease of access. Rather than being locked down to a local storage device that may not even is portable, you can use the power of the Internet to access or even stream in some circumstances from any location that has an Internet connection. Companies do not need to install physical storage devices in their own datacenter or offices, but the fact that storage has to be placed anywhere stays the same (maybe localization costs are lower in offshore locations) [9].Disadvantages of cloud storage are first , performance of cloud data storage is limited by bandwidth. Internet and WAN speeds are typically 10 to 100 times slower than LAN speeds. For example, accessing a typical file on a LAN takes 1 second, accessing the same file in cloud data storage may take 10 to 100 seconds. While consumers are used to slow internet downloads, they aren't accustomed to waiting long periods of time for a document or spreadsheet to load. Second, Availability of cloud data storage is a serious issue. Cloud data storage relies on network connectivity between the LAN and the cloud data storage provider. Network connectivity can be affected by any number of issues including global networks disruptions, solar flares, severed underground cables and satellite damage. Cloud data storage has many more points of failure and is not resilient to network outages. Network outages mean the cloud data storage is completely unavailable. Cloud data storage providers use proprietary networking protocols often not compatible with normal file serving

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 2, March 2013

on the LAN. Accessing cloud data storage often involves ad hoc programs to be created to bridge the difference in protocols [10].


**Fig 8 Cloud storage**

### B. *STRATERGIES FOR DELETING /REMOVING UNWANTED DATA FROM DATABASE*

Databases grow over time, and it is critical to have a strategy to deal with this. For every piece of data you need to determine how long you are going to keep it. Once you determine the lifespan of the data you will have to put in place a means of removing it. It is a sad fact that this is often neglected until a problem arises that demands it be dealt with. Company deletes its data at different interval of time and parameters for deleting data are also different. Below given some example, "How long we store data – Your data is stored in our databases for at most two months. After this period of two months, we can continue to store your data in anonymized way for a period of two years." In above Example Company store its user's data for two months in its database. After that they store user's data in anonymized way only for two years. There are two basic types of data that we talk about when we are talking about retention: Temporary and Semi permanent.

### C. *Temporary data*

Any data which is only necessary for a short period of time is considered temporary data. These tables will typically have a large number of inserts that are only valid for a narrow window such as session tables. Query performance will degrade over time if you fail to prune this table regularly. It is not uncommon for the developer to write logic which will invalidate a session after a specific period of time, but neglect to delete it. This is a mistake seasoned developers have made, so be sure to take it seriously.

**Strategies for Managing Temporary Data**

- Delete all expired entries
- Delete all entries that are older than one hour. These deletes can either be done from a separate utility, or embedded inside of the application.

### D. *Semi permanent*

Any data which you intend to keep for a long period of time but not permanently is considered semi permanent data. You will need to determine how long you will be keeping the data and then choose a strategy for managing it.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 2, March 2013

**Strategies for Managing Semi permanent Data**

Small data sets can be managed easily by simply deleting based on the age of the record and larger data sets should have greater thought put into the management strategy.

- Deleting old Records: deleting records from a database is a DML operation that can be quite time consuming. Very large deletes can block your database on reads and prevent your application from operating normally. There are a couple simple techniques you can do to minimize this impact however. First only perform the deletes during off-peak hours. Second put a limit on the delete call and call it multiple times. Those two approaches can minimize the negative impact of the large deletes. It is worth noting that you must properly index your time fields or your delete options will be painfully slow.
- Partition: if your database supports partitioning you can manage it via DDL operations that are near instantaneous. If you partitioned your database by months you could apply your retention policy by adding and removing months at a time. This action is very rapid consuming very little disk i/o. Partitions are an elegant solution as long as the version of database you are running supports them [11].

### E. Another way to save disk space is Auto shrink:

To compact a database, you create a new database, and then copy all objects from the source database to the new database. Typically, compacting is not initiated automatically. Automatically adjusting the size of a database file is called Auto Shrink. This technique uses almost no processor time and memory, making it especially suited to handheld devices and mobile database products. The Auto shrink technique moves pages within a file so that all the empty or unallocated pages are contiguously positioned at the end of the file. The empty pages are then truncated. Truncated pages are then available for the database file system to use. Returning the truncated pages to the database file system increases file system space [12].

## VII. ADVANCED DATA STORAGE DEVICES

### A. *Holographic Data Storage*

It is a new kind of technology that substantially helps in high-capacity data storage. This technology enables a device to record and read data at a million times faster rate than that in today's optical and magnetic drives. Fundamentally, it works on the 3D data storage concept. The data processing in this holographic concept relates to accessing the whole of the storage medium rather than just its surface. Data is stored on the disk using two rays intersecting each other. In this way, different data can be stored on the same photosensitive cells that hold data. So there can be data overlapping and hologram creation, wherein the data can be accessed using a beam equivalent to the hologram. At present, there are Holographic Versatile Discs (HVDs) and Holographic Versatile Cards (HVCs) developed, but are not yet commercially announced [13].

### B. *Memory Stick XC*

SanDisk and Sony announced the Memory Stick Format Series for Extended High Capacity. The Memory Stick XC has a maximum 2 TB capacity, 64 times larger than that of the Memory Stick PRO which is limited to 32 GB. XC series has the same form factors as PRO series, and supports Magic Gate content protection technology as well as Access Control function as PRO series does. In line with the rest of the industry, the XC series uses the newer exFAT file system due to size and formatting limitations of FAT/FAT16/FAT32 file systems used in the PRO series. A maximum transfer speed of 480 Mbit/s (60 Mbyte/s) is achieved through 8-bit parallel data transfer [14].

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their careful reading of this paper and for their helpful comments.

## REFERENCES

[1] http://www.emc.com/leadership/programs/digital-universe.htm (Accessed 2013-03-27)

[2] http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf (Accessed 2013-03-27)

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 2, Issue 2, March 2013

[3] http://www.cloudtweaks.com/2010/09/cloud-computing-storage-and-the-exponential-growth-of-data(Accessed 2012-12-12).

[4] http://books.google.co.in/books?id=sCCfRAj3aCgC&pg=PP15&lpg=PP15&dq=Information+storage+is+a+central+pillar+of+ information+technology.+A+large+quantity+of (Accessed 2013-03-27).

[5] http://www-01.ibm.com/software/data/bigdata/ (Accessed 2013-03-27).

[6] http://www.regentsprep.org/Regents/math/ALGEBRA/AE7/ExpDecayL.htm (Accessed 2013-03-27).

[7] https://docs.google.com/viewer?a=v&q=cache:Vaylz Ktu_8cJ:www.buffalotech.com/ (Accessed 2013-03-27).

[8] http://www.santrainingblog.com/2010/01/network-attached-storage-nas/ (Accessed 2013-03-27).

[9] http://en.wikipedia.org/wiki/Cloud_storage (Accessed 2013-03-27).

[10] http://www.clouddrive.com.au/download/ www.clouddrive.com.au-WhitePaper.pdf (Accessed 2013-03-27).

[11] http://commons.oreilly.com/wiki/index.php/Design_a_ Strategy_for_Removing Unwanted Data (Accessed 2013-03-27)

[12] http://msdn.microsoft.com/en-us/library/ms172411 (v=sql.105).aspx (Accessed 2013-03-27).

[13] http://www.buzzle.com/articles/holographic-datastorage html (Accessed 2013-03-27).

[14] http://en.wikipedia.org/wiki/Memory_Stick (Accessed 2013-03-27).

**AUTHOR BIOGRAPHY**

Ms. Pradnya Eknath Gaonkar is Assistant Professor in department of computer Engineering of Xavier Institute of Engineering, Mumbai. She is B.E in Computer and pursuing M.E. in Information Technology from Vidyalankar Institute of Technology, Mumbai.

Prof. Sachin Bojewar is Assistant Professor in department of Computer Engineering of Vidyalankar Institute of Technology, Mumbai. He is B.E. and M.Tech in Computer Engineering .He is also pursuing Ph .D in Computer Engineering. He has 17 Years of teaching experience.

Mr. Jayesh Ajit Das is currently working as Business Intelligence Consultant at Maia Intelligence, Navi-Mumbai. He is B.E. in Computer Engineering. He has 2 years of experience in Business Intelligence.