Storage Requirement Forecasting Analysis Model for Storage Area Networks

P.Mahalingam
System Administrator,
Caledonian College of
Engineering, Muscat,
Sultanate of Oman

Dr.N.Jayaprakash
Professor and Dean, School of
Computer Applications
Professional Group of Institutions,
Coimbatore, India

Dr.S.Karthikeyan
Asst. Professor, Dept. of IT,
Applied College of Arts and
Science, Sohar,
Sultanate of Oman

ABSTRACT

Information management is an important area to concentrate for the business continuity of an organization. Organization should have a plan to store, retrieve and maintain its valuable information to meet the business demands. Storage Area Network (SAN) is a high performance network available to meet the enterprise storage solution. Since, SAN uses Fibre Channel (FC) as a transporting medium, it is considered to be fast, reliable and an optimum solution to handle the explosive growth of digital contents due to internet, application and use of modern gadgets. Various storage models are available for enterprise storage and choosing the relevant storage as per the business requirement needs careful analysis of the present and future storage consumption inside the organization. The SAN implementation in an organization involves many individual component analyses including storage requirement/capacity planning to handle the business information's effectively and for the future. This research paper discusses a part of SAN implementation process, information generation, handling and information growth so that the SAN can be designed to meet the requirements. The aim of this paper is to analyze capacity of SAN as per the data generation and forecasting the data growth. The results were obtained for next five years in an organization using linear forecasting model to implement SAN as the enterprise storage solution.

General Terms

Enterprise Storage Design Analysis, Storage Area Networks Implementation, Storage Design Capacity Planning Analysis, Forecasting of Future Storage Usage.

Keywords

Information Management, Information Growth, Storage Area Networks, SAN Design, Storage, Network, Capacity Planning, Linear Forecasting,

1. INTRODUCTION

Business automation and modern gadgets increased the digital contents extensively. Internet is another reason where the digital universe got bigger as was never before. Organization's storage requirement is growing three fold every year [1] and this indicates that the digital universe increasing its size by 75% annually. Over the last several years, the percentage of data increase is accelerated and prompting a drive toward handling this differently by considering its performance, production, access and delivery [3]. The first hard disk drives (HDD) introduced by IBM Corporation in 1956 [2] has provided the storage industry a platform on which many devices was built to handle the data. Due to the invention of Hard disk drives, the storage technologies have changed and have improved tremendously up to its latest Storage Area Networks (SAN). There are three storage models commonly used in the storage industry to provide many improved, customized storage devices manufactured by various companies such as Seagate, Hitachi, Sony, IBM, SUN, EMC and HP. The three common storages models are i) Direct Attached Storage (DAS), ii) Network Attached Storage (NAS) and iii) Storage Area Networks (SAN). Direct Attached Storage (DAS) – wherein storages are connected directly to the workstation or desktop. Network Attached Storage (NAS) – where in storage is connected to a dedicated server (normally called file server) and linked with Local Area Network (LAN) for shared access. Storage Area Networks (SAN) - an isolated high speed network usually interconnected with Fibre Channel (FC) wherein heterogeneous storage devices (HDD, Tape, Laser Disks) are connected externally with block level data transfer. SAN environment should be designed based on the business process needs and how it is offered by different manufacturers. SAN normally provides high speed data transfers from Server-to-storage, Server-to-server and Storage-to-storage. SANs allow data to be moved without server intervention, freeing up server processor cycles for other activities like application processing. Some examples of this communication include a disk device backing up its data to a tape device without server intervention, or remote device mirroring across the SAN. SAN provides maximum benefits from the available storage models with secured data access, high data availability, scalable, reliable and high speed data transfer makes. If an organization wants to implement the SAN as storage solution, it is necessary to collect the information about the requirement of storage space needed [4] for the present and future storage utilization in the initial stage of SAN design and implementation. The storage requirement analysis involves an account of information's generated by users, business and external customers and expected data generation. The organization should carefully consider collecting information such as document creation, database, internet content, emails and audio/video contents to provide the effective data management. The storage forecast gives an idea of how much people need to manage the enterprise storage. Forecasting is the estimation of the value of a variable (or set of variables) at some future point in time. A forecasting exercise is usually carried out in order to provide an aid to decision-making and in planning the future [17]. Forecasting normally helps the SAN designers to build the SAN with the required storage libraries.

2. LITERATURE REVIEW

2.1 Storage and Information Growth

Computer Storages are the backbone of modern businesses wherein the information's to be managed efficiently in order to compete with the business. Computer data storage, often called storage or memory, refers to computer components, devices, and recording media that retain digital data used for computing for some interval of time [15]. Many types of computer storages are used in the industry to store and retrieve the data. One of the best storage options available to handle the huge volume of information for the today's business is Storage Area Networks or SAN [5]. The substantial information growth comes more and more when industries such as banks, news agencies, medical industries, academic portals, production areas and digital libraries are automated and all the information's is started stored in computers and databases for the quicker access [6]. The famous video sharing portal, YouTube which does not exist few years back, hosts 100 million video streams per day and more than billion songs are uploaded or downloaded [7].
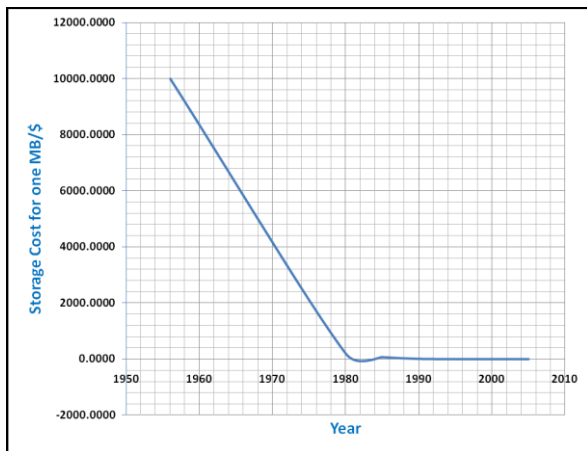

Fig 1: Yearly Improvement in Rate per MB Storage [1]

According to Storage Networking Industry Association (SNIA), the SAN is defined as a network whose primary purpose is the transfer of data between computer systems and storage element and among storage elements. The major benefits of SAN includes increased disk usage, improved sharing among users, secured transmission of data over the network, centralized backup, storage consolidation, reduced TCO, enhanced manageability, better disaster recovery and high availability. Network attached storage (mostly NAS) will play a more significant role as content sharing becomes more common, network connectivity and management is simplified and more and more consumer devices are built with network connectivity [19]. The storage technologies are developed on par with the growth of digital information and the storage cost reduces is direct proportion to its capacity "Fig. 1". The storage world kick-started towards high capacity into terabytes where the prices started to drop to less than a hundred dollar for 100 GB hard disk

2.2 Storage *vs*. Information

The enormous increase in digital information poses a threat to the industries where information's are vital assets for business continuity. A large amount of storage space usage involves maintenance of many storage devices such as magnetic tapes, hard disk drives, optical storage devices etc.

Another major storage space consuming internet application is digital libraries hosted in the internet for shared access. Many universities and educational institutions started hosting their books, journals, course materials, dissertations and research papers in the form of digital contents for access by millions around the world through internet. The sudden rise in social network usage such as Orkut, face book, personal blogs and web mails storage space are major sources of digital contents. The sources of digital content creations widened due to the modern devices "Table 1" and internet technologies such as web contents, emails, ecommerce information, online trading and much more. The application specific data, mobile computing, healthcare, banking data, ATM, digital printing, digital libraries, online books and journals also increased the digital universe into high volume data container. The world information growth versus storage devices to handle this large volume of digital information is given in "Fig 2".

As information is growing on one side, the storage industry is being pumped with a wide range of storage devices and technologies to handle the explosive growth of information. The information growth has put the storage professionals in to high risk for managing and securing the stored information. "Table 1" lists various sources of digital content creations and individual storage devices available.
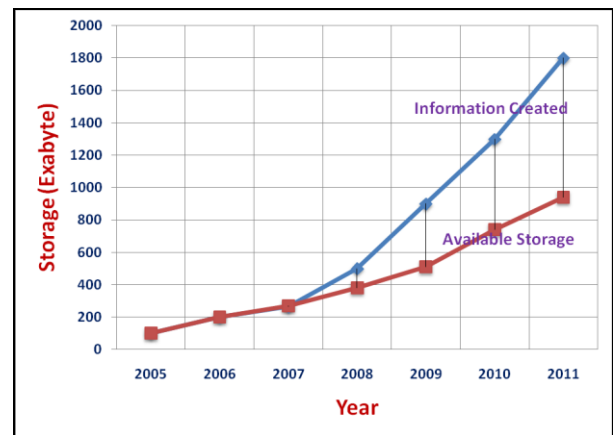

Fig 2: Information Growth vs Storage Devices [1]

It is evident that, the modern gadgets and internet applications took major share of sources of digital information creation. The storage technologies are also developed on par with this information growth to handle the stored information effectively and efficiently. Repetition of information stored in the storage devices is another area which occupies a major storage space. Information created by the user is stored in desktop, server, backup external hard drive and in emails which occupies the storage space.

Table 1. Various sources of information and storage devices

INFORMATION'S AND STORAGES
Capturing/Creations
High end camera/Digital cameras, Camcorders/Camera phones/ web cam, Surveillance/Scanners
Multifunction peripherals, OCR/bar code reader
Medical imaging, Digital TV & video, Graphic workstation, Digital voice capture, Landline telephone/VoIP, Mobile phone
Data Creations
PC Applications & Databases, Office applications, Email & Video teleconference , Chat/ IM/Smart phone/ PDA, Terminals/ ATM/Specialized PCs, Industrial machines/Home appliances, Production equipment/RFID/ sensors, Smart cards/ Video games, MP3 player/ Ipod/SMS/MMS/GPs, Servers/Business firms/Application development, IT infrastructure/Technical
Data Creations
HDD, Optical , Tape , NV flash memory, IC, Laser disc

3. DATA STORAGE USAGE ANALYSIS

The storage requirement analysis gives the overall picture of how much storage space is used presently and how much storage is required for the future. Based on the storage requirement analysis, the storage technology will be selected and used. The storage requirement will be based on how much information created from the business or users. In an academic environment, information is generated from many services, which should be kept for future usage. For example, the academic staff will generate various course materials, question papers, handouts, presentations, emails and web contents. Blackboard is another storage consuming platform wherein the course materials will be uploaded for the students access. The course materials will be simple word documents to sophisticated audio/video contents. The course materials will normally be prepared by the academic staff.

The basis of digital content growth of an organization is how much content is created on several of its activities. The major areas of information growth comes in an organization from the email and websites where enormous information are downloaded and uploaded in the form of audio, video and photos. As the world is progressing in creating the digital contents very rapidly, this study aimed to know the organizations information growth so that it will also help in finding the capacity for the SAN implementation.

The SAN design analysis environment selected for this work was a premier university college running with 900 workstations and 16 servers. The college has over 3000 students and 250 staff including academic and administrative personnel. Data was collected for the purpose of storage requirement analysis. The SAN design as per the forecasting for the next 5 years will be done based on the data collected for the present utilization of storage is the central file server, web server and email servers.

Table 2. Storage Requirement Data Collection Form

Storage Requirement Analysis – Data Collection Sheet	
Name:	Academic Staff: □ Administrative Staff: □ Student: □
Access to Central File Server	Yes/No
Web Access	Yes/No
Email Access	Yes/No
Present Storage Usage (PC Storage & Server Storage)	_____ GB
Documents Creations	_____ GB
Web Contents Access	_____ GB
Email Send/Receive	_____ GB
Audio/Video Downloaded/Stored	_____ GB
Software's/Applications Installed	_____ GB
Date Recorded: _____	Recorded by: _____

The data was collected among the staff and students of the environment where the SAN is going to be implemented. Random sampling method was adapted for data collection. The sample size included 20 staff drawn from academic and administration and 20 students representing various levels of study. A questionnaire (see "Table 2") was used to collect the information about the storage usage for the various day-to-day work including document creations, audio/video, course materials and application software's used. After the data collection, the data were analyzed using the statistical tool for linear forecasting to find out the total storage space required for next five years.

The statistical software NCSS Pass 2008 [16] was used for forecasting SAN capacity analysis. The forecast analysis was done as per the data collected and the known values, years and storage space used was y and x. The formula, used wherein x and y is replaced with year and sample means of storage. The R squared value also called as coefficient of determination shown in the forecasting graph varies from 0 to 1, to reveal the forecasting value closer to corresponding trends of the actual data

$$a = \bar{y} - b\bar{x} \qquad b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

Initially, the data was collected using questionnaire on weekly basis and summarized on monthly mean and further summarized for year-wise data.

4. STORAGE CAPACITY FORECASTING

Determining the storage capacity for the business requirement is one of the major analyses in SAN deployment. The data should be carefully collected which reflects the actual requirement of the business [18]. The data was collected as per the questionnaire designed and given in "Table. 3". The samples collected were divided in to documents created by users using application software's such as Word, Excel, PowerPoint and

other engineering applications, emails usage with attachments, web access contents including the downloading of documents and web pages. Further, the audio and video created and uploaded from various sources such as emails, mobile phones, cameras and web sites were recorded separately. The data created by the users were analyzed and is presented as shown in Fig. 3 to 8. The analysis shows the total storage space used and storage requirement for the next 5 years.
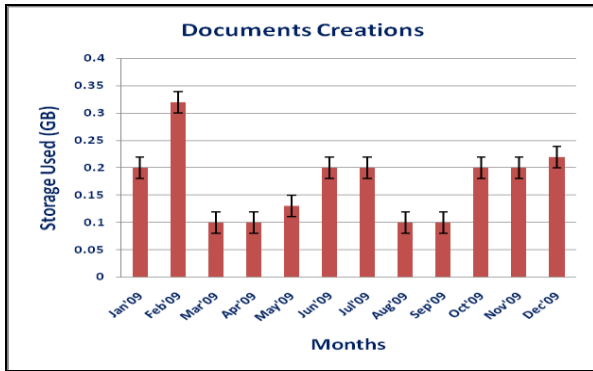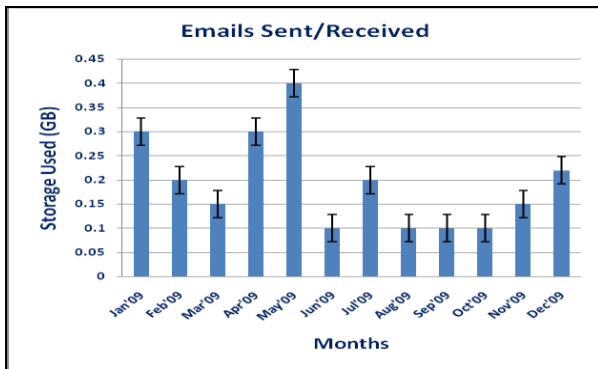

Fig 3: Information Growth - Documents
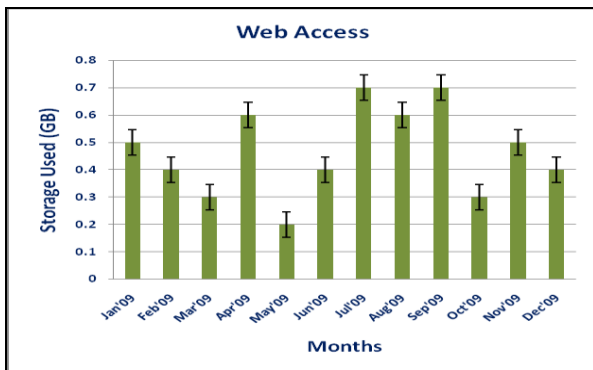

Fig 4: Information Growth - Emails

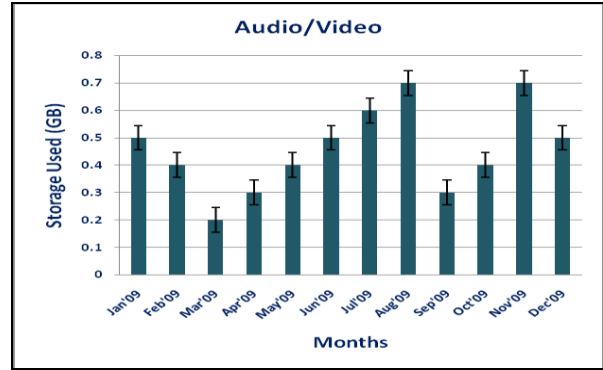**Fig 5: Information Growth – Web Access**
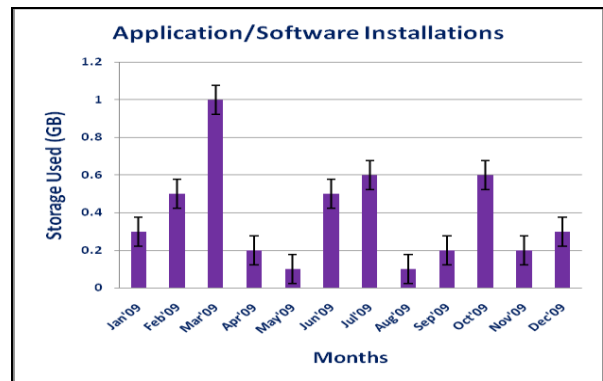

Fig 6: Information Audio/Video

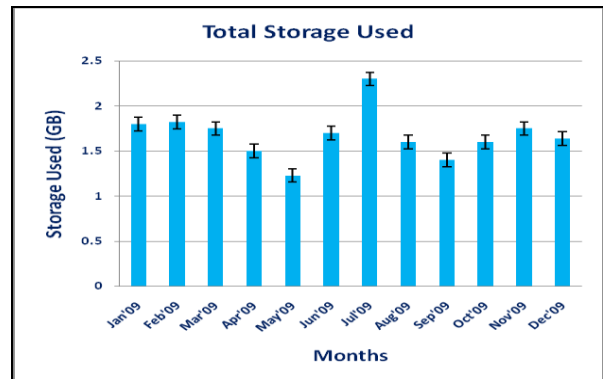**Fig 7: Information Growth – Applications**

Fig 8: Information Growth – Storage Used

5. RESULT AND ANALYSIS

The results are considered as average trend in information creation throughout the year. The standard error 5% is an acceptable limit. The linear model of forecasting of the storage requirement analysis gives r-squared value equal to 1 which indicates that the data growth will be linear growth, the vertical growth of information's among the users. The data collected from various users are complied and summed-up to month's data. The data analyzed further for the mean average into yearlong storage space used. The yearlong data compiled and analyzed for forecasting for next 5 years is given in "Fig. 5". The information growth is found to be linear and almost r-

squared equal to 1, which indicates the data, was very close to the actual requirement. The summary also gives an idea of deciding the SAN capacity for the next 5 years. Based on the data collected and analyzed using the forecasting analysis, the growth of information predicted for the next year is around is 13 TB. The proposed SAN design analysis and capacity model forecasted 13 TB as per the data collected and analyzed. The forecasted SAN capacity should be increased a little higher, so that any unexpected increase in information generation or storage requirement will also be handled effectively.

6. CONCLUSION

Information are the assets of an organization and information management is the critical area where efforts are to be made for its effectiveness. SAN provides dynamic storage solution for the enterprise business requirements and needs careful analysis before design and its deployment. The growth of business is based on the effective management of its valuable information in today's information era. High availability of data to the users, who needs it, is the primary requirement of a business and SAN meets this requirement effectively. In order to implement the SAN, the requirement analysis is to be carried out and the SAN capacity planning is to be done to make sure that the proposed SAN solution never runs out of the space for the business requirements. This study explores the possibility of measuring for storage usage of the present and proposed a model forecast the future storage requirement. The methodology designed in this paper try to reflect the actual storage requirement based on the collected data. This paper also affirms the three fold information growth as found in the literatures. The results from this study will be used for the effective design and deployment of Storage Area Networks

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Kevin Craine. The Growth of Digital Information, April, 2001. Available: http://www.tdan.com/view-articles/4917

[2] Morries RJT & Truskowski BJ. The evolution of Storage Systems, IBM Systems Journal, Volume 42, Issue 2, Proquest Science Journal, 2003. pp.205-217

[3] Jamie Gruener, Research note: Compliance and digital-content growth drive content-storage market, The Yankee Group , May 20, 2003

[4] Odysseas I. Pentakalos, Daniel A. Menascé, Milton Halem, Yelena Yesha, Analytical Performance Modeling of Hierarchical Mass Storage Systems, IEEE Transactions on Computers, vol. 46, no. 10, pp. 1103-1118, October, 1997.

[5] G. A. Gibson, D. F. Nagle, K. Amiri, J. Butler, F. W. Chang, H. Gobioff, C. Hardin, E. Riedel, D. Rochberg, and J. Zelenka, "A cost-effective, high-bandwidth storage architecture," in ASPLOS-VIII: Proc. of 18th inter. conference on Architectural support for programming languages and operating systems, vol. 32, no. 5. New York, NY, USA: ACM Press, December 1998, pp. 92-103

[6] Gank and Rochester. Pirates of the digital millennium. FT Printice Hall, 2005, pg 175.

[7] Jhon F Gantz. The Diverse and Exploding Digital Universe – An IDC White Paper, Sponsored by IDC, March 2008.

[8] Aziz, M.H.Ong Con Nie, Jesse Chan Mei Yam & Lee Chang Wei. TCO reduction. Communications, 2003. APCC 2003. The 9th Asia-Pacific Conference. Volume: 3, On page(s): 1147- 1151 Vol.3. ISBN: 0-7803-8114-9, 2003.

[9] D. R. Avresky, V. Shurbanov, R. Horst, W. Watson, L. Young and D. Jewett. Performance Modeling of ServerNet™ SAN Topologies. ISSN. 0920-8542, Springer Netherlands, The Journal of Supercomputing, Computer Science, pp.19-37, December, 2004

[10] Storage Area Networks Design Reference Guide, Hewlett-Packard, pp.26. Sep 2003. Available: http://www.ost-india.com/services/san/understanding%20sans.pdf

[11] Jiwu Shu, Bigang Li, Weimin Zheng, "Design and Implementation of an SAN System Based on the Fiber Channel Protocol," IEEE Transactions on Computers, pp. 439-448, April, 2005

[12] Ng, J. M. and Susilo, R. N. Controlled shared backup strategy based on channel availability. Int. J. Comput. Appl. Technol. 24, 1 (Jun.2005), 25-32. DOI= http://dx.doi.org/10.1504/IJCAT.2005.007202

[13] Erik Riedel, Mahesh Kallahalla, and Ram Swaminathan. A Framework for Evaluating Storage System Security. Proc. of the 1st Conference on File and Storage Technologies (FAST'02), Monterey, CA, January, 2002.

[14] P.Mahalingam, N.Jayaprakash & S.Karthikeyan. Enhanced Data Security Framework for Storage Area Networks. Proc. Of the 2nd International Conference on Environmental and Computer Science (ICECS 2009, IEEE Explore), Dubai, UAE, pp105-110, 2009.

[15] Computer Data Storage, Wikipedia. Available at http://en.wikipedia.org/wiki/Computer_data_storage

[16] Forecasting and Time Series Analysis using NCSS, NCSS, USA. http://www.ncss.com/4cast.html

[17] JE Beasley. Introduction to Forecasting, OR-Notes, http://people.brunel.ac.uk/~mastjjb/jeb/or/forecast.html

[18] Paul Ross. Design Considerations in Enterprise Storage Networks.

[19] Industry Trend or Event. Computer Technology Review, Nov, 2000.