International Journal of
Electronic Commerce Studies
Vol.2, No.1 , pp.77-86, 2011

GROWTH OF DIGITAL CONTENTS AND INFORMATION MANAGEMENT USING STORAGE AREA NETWORKS

P. Mahalingam
Caledonian College of Engineering
mahalingam@caledonian.edu.om

N. Jayaprakash
Professional Group of Institutions
drnjprakash@gmail.com

S. Karthikeyan
Applied College of Arts and Science
skaarthi@gmail.com

ABSTRACT

The Internet has changed information handling and escalation of digital contents in recent years, but appropriate storage technology is needed for effective management of information which is critical to businesses today. The rapid growth of digital contents requires a technology that delivers high availability, scalability, and reliability. Storage Area Network (SAN) is one promising solution providing high speed data transfer with many other storage services, backboned with a high speed Fibre Channel (FC). SAN design and implementation requires careful analysis before SAN can be used as an enterprise storage solution. This research paper discuss information growth, information handling, and various storage technologies available today, in addition to financial justifications, backup procedures, disaster management and security measures. This paper also recommends an efficient design analysis considering various factors for SAN implementations to meet the customer demands of organizations.

Keywords: Information Growth, Information Management, Information Security, Storage Area Networks, Total Cost Ownership, TCO, Return on Investment, ROI.

1. INTRODUCTION

The explosive growth of digital content in the form of text, pictures, audio, and video is extraordinary due to the wide use of the Internet and modern digital gadgets, and the sources of digital content increase each day. The storage requirements of organizations are growing three fold every year [1, 2], indicating the digital universe is annually increasing its size by 75%. An interesting fact to note is that the growth of digital information in the last 30 years is equivalent to information growth in the previous 5000 years of the entire history of civilization, thanks to the invention of Internet technologies.

Storage technologies are the backbone of modern business and need constant update and upgrade in order to meet customer demands. The first mass storage device to store data was introduced by IBM corporation in 1956[3], called hard disk drives (HDD), they provided the platform on which the storage system industry has been built. Due to the invention of hard disk drives, storage technologies have changed their face and improved tremendously up to the latest Storage Area Networks. Information management deals with the collection of information from various sources, storage, and access by various users[4]. It is important to financially justify SAN implementations by evaluating Total Cost Ownership (TCO) and Return On Investment (ROI) relative to expenses, as this impacts business profit. Disaster management is one of the major areas in information management. During the information process, if data is lost due to unforeseen situations, the disaster plan should provide for recovery of the original information. Data integrity is very important to provide accurate and reliable information to customers, and it will be achieved through application of appropriate security techniques[5] for the stored data.

2. LITERATURE REVIEW

2.1 Information Growth

Data is often viewed as the lowest level of abstraction from which information and knowledge are derived. There are a wide variety of digital information sources which are increasing in due to the invention of modern gadgets. Some of the sources of information are: mobile, digital camera/video camera, documents, emails, Web contents, audio, video, application specific data, E-commerce information, online trading, web page contents, mobile computing, IP telephony, healthcare, banking data, digital printing, digital archiving, digital TV, and home appliance system based

digital data. Storage architects predict[1, 6] that information growth will double in the next three to five years (see Figure 1). As information volume continues to grow, the storage industry is pumped with a wide range of storage devices such as direct attached storage (DAS), network attached storage (NAS) and storage area networks (SAN). The enormous increase in digital information poses a threat to industries where digital information is a vital business community asset. Digital information growth can be illustrated by just one case. The famous video sharing portal, Youtube.com, which did not exist a few years ago, hosts 100 million video streams per day [7]. More than a billion songs are uploaded or downloaded over the Internet each day.

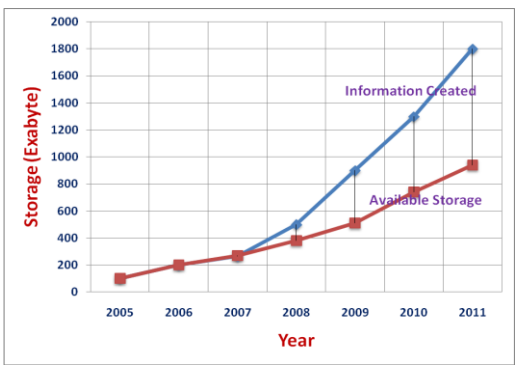

Figure 1. Information Growth vs. Available Storage[1]

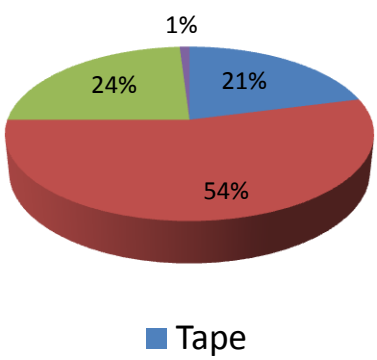

Figure 2. World-wide Usage of Different Storage[2]

2.3 Issues due to Information Growth

The growth of the digital universe creates problems in the areas of storage administration and maintenance and creates burdens to the people who are involved in maintenance, backup, and restore activities[4]. Security is another issue needing much time and concentration to protect stored data. Many people are involved worldwide in developing methods to safeguard the integrity of data. The following are the some of the key challenges faced by storage professionals and managers: (1) High storage costs (2) Designing and deploying storage technology (3) Managing backup/recovery (4) Managing security for large data storage (5) Managing data availability (6) Shortage of storage professionals (7) Selecting the right storage model and (8) Lack of literature for new storage models. There is a gap in the storage available versus the information created so far (see Figure 1). This gap may widen if the information continues to grow in this manner. Keeping this in mind, the storage industry has already started developing relevant technologies to handle this volume of information.

2.3 Storage and Storage Models

Computer data storage, often called storage or memory, refers to computer components, devices, and recording media that retain digital computing data for some interval of time. Many types of computer storage are used in industry to store and retrieve data, as shown in "Fig. 2". One of the best storage options available to handle the huge volume of information for today's businesses is Storage Area Networks or SAN[8, 9]. The major benefits of SAN include 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. One of the unique characteristics which take SAN to the top of storage solutions is its use of storage in an external environment[10]. The world-wide storage usage of various storage devices is given in Figure 2.

3. INFORMATION MANAGEMENT THROUGH SAN

3.1 Storage Area Networks

There are three storage models[11] commonly used in the storage industry to provide the many improved, customized storage devices manufactured by various companies. The three common storage models are (see Figure 3), Direct Attached Storage (DAS), Network Attached Storage (NAS), and Storage Area Networks (SAN). Direct attached storage uses,

storage connected directly to the workstation or desktop. Network attached storage uses, storage that is connected to a dedicated server and linked with a local area network (LAN) for shared access. NAS network users centrally share storage, with limited scalability. Storage area networks use an isolated high speed network, usually interconnected with fibre channel. Heterogeneous storage devices (HDD, Tape, Laser Disks) are connected externally with block level data transfer. SAN normally provides high speed data transfers in the following three[12] ways: *Server-to-storage*, *Server-to-server* and *Storage-to-storage*. In a business environment, effective information management requires various factors to be considered before SAN design and deployment:

a) Total Cost Ownership and Return on Investment

b) Data Location, Storage Libraries, and Inter-connectivity

c) Backup/Restore and Disaster Management

d) Data Availability and Manageability

e) SAN Security

f) Performance and Monitoring

a) Total Cost Ownership (TCO) and Return On Investment (ROI)

In TCO analysis, the various costs were calculated as per business requirements. There are two types of costs, hard and soft. Hard costs include total purchase cost, total administration costs, and total utilization costs, and soft costs include storage availability and cost per downtime hour [13]. The installation costs are also included along with storage personnel salary. The ROI is calculated based on the amount spent, calculated from the TCO. In order to get the perfect ROI for SAN deployment[14], three models should be analyzed: Payback Period (PP), Internal Rate of Return (IRR), and Net Present Value (NPV). The simplest way to analyze the ROI is given in the following formula:

ROI = (Sum of all Income – Sum of all Investments) / Sum of Investments

b) Data Location, Storage Libraries, and Interconnectivity

Storage devices and server deployment should be based on the specific requirements for data location. With regard to the SAN topology design, location involves the SAN storage system placement relative to servers accessing the storage. The interconnectivity can be done through any one of the three types of relationships:[12], one-to-one, many-to-one, and

many-to-many. The storage libraries should be selected as per available capacity, based on the storage requirement analysis. The devices should have redundancy so failures can be handled in a better way. One example is RAID technology where the mirroring is done at the disk level to prevent heavy data loss. Interconnectivity is based on the data location and distance from access workstations. Fibre channel gives block level access[15] so that the data transfer will be high speed which is the primary factor of adapting SAN as enterprise storage.

c) Backup/Restore and Disaster Management

SAN provides centralized backup activity carried out in a secured environment; SAN backup tends to be highly reliable and redundant. The backup strategies[16] need to be developed with customized guidelines which includes periodic of backup, backup method and many more. The disaster recovery management (DRM) system is an important part of safeguarding business data against any unexpected disasters. DRM involves processes such as (1) analysis of the business requirement (2) formulating the DRM team (3) defining the DRM policy and procedure (4) developing the detailed DRM plan and activities (5) implementing and updating the DRM plan.

d) Data Availability and Manageability

Data availability refers to how reliable stored data is for day-to-day operations. The preventive maintenance of all SAN equipments keeps it alive for long times and reduces the failures of devices. Administrators must carve up the storage space into segments that are only accessible to specific users[11]. SAN management should be done by specialized personnel who are skilled in storage activities; SAN personnel should be involved from the beginning of SAN implementation including installation and configuration.

e) SAN Security

SAN security is the group of parameters and settings that makes storage resources available to authorized users and trusted networks and unavailable for others. Storage Area Networks security comprises many methods and approaches that are linked together to achieve the highest security of stored data[17]. Storage Area Networks is one effective technology for organizations that need confidentiality, integrity, scalability, availability, and high performance storage solutions. Since SAN holds mission critical data which is very vulnerable to attackers, it needs multi-factor considerations to address security issues[18]. Effective storage management should implement five basic areas of security in each level of SAN, namely, Storage Array Volume Access Control, Volume Access Control on the HOST, Device Configuration Access Control, Storage Management

Software Access Control and Proactive Detection of Access Violation, and Auditing and Logging. The five broad categories are further individualized into Access Control (Zoning & LUN masking, Intrusion Detection System (IDS), Cryptography (CFS, SFS & EFS), Authentication and Authorization, Fibre Channel Security, and Security by SAN Management Software. When designing a SAN security framework the many factors described (see Figure 4) should be considered.

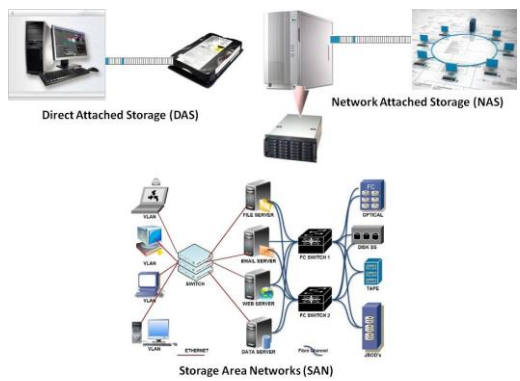

Figure 3. Three Storage Models

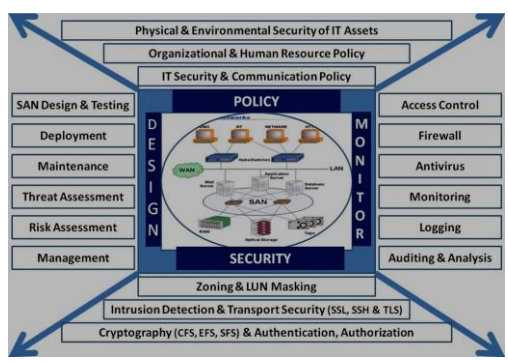

Figure 4. SAN Security Framework

f) Performance and Monitoring

In general, performance refers to the functioning and effectiveness of a particular entity. In SAN deployment, performance analysis involves the analysis of Bandwidth (storage bandwidth, disk, SCSI or FC), Throughput

(the amount of work performed by a component or system over unit time[19, 20]), Delay (wait time induced by contention within a system), Latency –(amount of time between the initiation of an action and the actual start), Response time (time it takes to finish a given storage operation), Scalability (ability of a SAN to grow without adversely affecting its services), Reliability (the degree to which a given computing component produces consistent results), and Availability (the degree to which a SAN keeps running within acceptable performance limits and without any unrecoverable failures). Monitoring records all the activities that are accessing data and the time this occurs, which helps give auditors security perspective.

4. CONCLUSION

Information plays a vital role in business continuity, and the Internet changes the face of information access. SAN is one promising solution to provide reliability, scalability, and high information availability for instant access. To increase productivity, meet customer demand, and be a leader in industry, any organization must ensure that its data is reliably stored and available all the time access is needed. SAN demands that a detailed study be conducted to analyze the design factors for maximum effectiveness. The data integrity of an organization, whether data-at-rest or data-at-transmission, solely relies on the security methods applied to protect it from internal and external attacks. Effective SAN design should always keep security methods in mind from the initial start of design process of SAN. An organization should develop internal and external data policies for the accessibility and use of its information by various personnel. Information management activities such as storing, retrieving, archiving, securing, and removing are much easier using SAN than any other storage solution.

REFERENCES

[1] J. Gantz, C. Chute, A. Manfrediz, S. Minton, D. Reinsel, W. Schlichting, and A. Toncheva, *The Diverse and Exploding Digital Universe. IDC*. Retrieved on July 12, 2010, from http://www.emc.com/colate/analyst-reports/diverse-exploding-digital-universe.pdf.

[2] Craine. Kevin, *The Growth of Digital Information*. Retrieved on July 12, 2010, from http://www.tdan.com/view-articles/4917.htm.

[3] R. J. Morris, and B. J. Truskowski, The Evolution of Storage Systems. *IBM Systems Journal*, 42(2), p205-217, 2003.

[4] James, R., *Principles of Information Management, Step Two Designs*. Retrieved on July 12, 2010, from http://www.steptwo.com.au/papers/kmc_effectiveim/index.htm.

[5] SafeNet, *Multi-Factor Authentication - Protecting Applications and Critical Data against Unauthorized Access*. Retrieved on February 29, 2010, from http://www.safenet-inc.com.

[6] Gank, and Rochester, *Pirates of the Digital Millennium*. FT: Prentice Hall, 2005.

[7] The Rise of YouTube: A Big Deal, *TechTalk, EMERGIC*. Retrieved on January 9, 2009, from http://archives.emergic.org/collections/tech_talk_the_rise_of_youtube.htm.

[8] Storage Area Networks, *Overview of SAN, StorageWorld*. Retrieved on February 29, 2010, from http://www.forbiddenhistory.co.uk/storageworld/tech/san/overview.htm.

[9] Taking the Mystery Out of SAN and NAS Convergence, *Hitachi Data Systems, Technical White Paper*. Retrieved on February 29, 2010, from http://www.strategic-reports.com/guides/SANNASCoexistenceWHP-149.html.

[10] Garth A. Gibson, David Nagle, Khalil Amiri, Jeff Butler, Fay W. Chang, Howard Gobioff, Charles Hardin, Erik Riedel, David Rochberg, and Jim Zelenka, *A Cost-Effective, High-Bandwidth Storage Architectur*. Proceedings of the 8th Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), p92-103, 1998.

[11] J. Bigelow. Stephen, *SAN Management Overview*. Retrieved on February 29, 2010, from http://searchstorage.techtarget.com/article/0,289142,sid5_gci1218232,00.html.

[12] Storage Area Networks Design Reference Guide, *Hewlett-Packard*. Retrieved on February 29, 2010, from http://www.ost-india.com/services/san/understanding%20sans.pdf.

[13] The Cost Benefits of a SAN, *An Analysis of Total Cost of Ownership of an iSCSI SAN, Fibre Channel SAN, and Direct Attached Storage, Comp. Technology Review*. Retrieved on February 29, 2010, from http://findarticles.com/p/articles/mi_m0BRZ/is_7_24/ai_n6185910.

[14] Executive Summary, *ROI–The Storage Perspective*, S*atyam Computers*. Retrieved on January 9, 2009, from http://www.scribd.com/doc/11541574/The-storage-perspective.htm.

[15] S. Jiwu, L. Bigang, Z. Weimin, and Zheng, Design and Implementation of a SAN System Based on the FC Protocol. *IEEE Transactions on Computers*, 54(4), p439-448, 2005.

[16] J. M. Ng, R.N. Susilo, Controlled Shared Backup Strategy Based on Channel Availability. *International Journal of Computer Application Technology*, 24(1), p25-32, 2005.
[17] E. Riedel, M. Kallahalla, and R. Swaminathan, A Framework for Evaluating Storage System Security. *Proceedings of the 1st USENIX Conference on File and Storage Technologies*, January 28-30, Monterey, CA, 2002.
[18] P. Mahalingam, N. Jayaprakash, and S. Karthikeyan, Enhanced Data Security Framework for Storage Area Networks. *Proceedings of the 2nd International Conference on Environmental and Computer Science (ICECS 2009)*, December 28-30, Dubai, UAE, 2009.
[19] D. R. Avresky, V. Shurbanov, R. Horst, W. Watson, L. Young, and D. Jewett, Performance Modeling of ServerNet™ SAN Topologies. *The Journal of Supercomputing*, 10(4), p19-37, 2004.
[20] O. I. Pentakalos, D. A. Menascé, M. Halem, and Y. Yesha, Analytical Performance Modeling of Hierarchical Mass Storage Systems. *IEEE Transactions on Computers*, 46(10), p1103-1118, 1997.