Storage Area Networks (SANs) in Business Environment

By

Mohd Razif bin Mat Rejab

Dissertation submitted in partial fulfillment of
the requirements for the
Bachelor of Technology (Hons)
(Information System)

JUNE 10, 2004

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

STORAGE AREA NETWORKS (SANs) IN BUSINESS ENVIRONMENT

Research on "STORAGE AREA NETWORKS and ARTIFICIAL INTELLIGENCE CONCEPT IN DECISION MAKING"

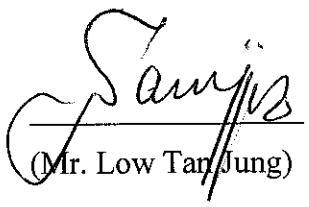
By

Mohd Razif bin Mat Rejab

Dissertation submitted in partial fulfillment of the requirements for the
BACHELOR Of TECHNOLOGY (Hons)
INFORMATION SYSTEM

10 JUNE 2004

Approved by,



(Mr. Low Tan Jung)

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the reference and acknowledgements, and that the original work contained herein has not been undertaken or done by unspecified sources or persons.

MOHD RAZIF BIN MAT REJAB

ABSTRACT

Storage Area Networks (SAN) in Business Environment is titled and initiated to design and implement Storage Area Networks architecture in the business operation. The project is divided into two terms, first is the research of Storage Area Networks and the second is system development on the Storage Area Networks Knowledge Management System. Research on the Storage Area Networks was based on the problem statement and objective of the project while the Storage Area Networks Knowledge Management System is the system in making decision to implement Storage Area Networks. The project will require a hybrid model for System Development Life Cycle (SDLC) methodology.

Reviews on the system will be made according to the SDLC and the objectives of the project. Artificial Intelligent module is used for the Storage Area Networks system to determine the best Storage Area Networks solution for the business. Research will be more on the implementation of the Storage Area Networks in the business based on the cost, availability and the architecture of the Storage Area Networks. Advantages of the Storage Area Networks and several criteria in the Storage Area Networks will be part of the Storage Area Networks research.

Storage Area Networks give the best solution for business as the database is an important asset for the business. Performance, availability, flexibility and scalability are the main subject in considering Storage Area Networks.

Keywords: Storage Area Networks, Knowledge Management System, hybrid model. System Development Life Cycle

ACKNOWLEDGEMENTS

First and foremost, Praise Be Upon Allah for His Mercy has given me the strength to complete my final year project and according to requirements.

I believed this a result of priceless contribution from many parties. I would like to express my sincere and hearties appreciation to:

- Mr. Low Tan Jung , final year project supervisor for her great help kindness, guidance and valuable ideas towards completing the project
- My parents and family for their moral support and financial support.
- My fellow friends, which help and give good advice in developing this system and other final year students of Information Technology and Information System, who have together strive thorough the semester completing the final year project.
- Manager and Executive of UTP IT Department, for their cooperation especially during interview.
- Finally thank you to everyone involve directly or indirectly with this project.

Your assistance and ideas are appreciated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND NOMENCLATURES

FYP – Final Year Project

SANs – Storage Area Networks

Win2K – Window 2000

MacOS – Macintosh Operating System

OLTP –Online Transaction Processing

LAN –Local Area Network

WAN – Wide Area Network

I/O – Input Output

SCSI – Small Computer System Interface

RAID – Redundant Arrays of Inexpensive Disks

SDLC – System Development Life Cycle

AI – Artificial Intelligent

GB – Giga bytes

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND OF STUDY

The storage market is at the beginning of a period of rapid change. Many of the underlying fundamentals in terms of how storage was viewed in relation to the rest of an enterprise's network are changing. Fiber Channel technology, which is replacing parallel SCSI in many environments is providing a much higher degree of connectivity in terms of both the number of devices that may be connected, and the distances between them. This change allows new storage configurations which take advantage of this connectivity to provide more scalable and flexible solutions. One recent concept which has received a lot of media and industry attention is the Storage Area Networks or SANs.

Storage Area Networks (SANs) are storage technology which is developed for today's storage efficiency in business environment. SANs are network of storage devices that are connected to each other and to a server, or cluster of servers, which act as an access point to the SAN (Figure 1.0*)*. The storage area network (SAN) provides a flexible, networked storage infrastructure that decouples storage devices from their respective servers. The implementation of SANs will offload traffic bandwidth from the production of LAN produced from servers and deliver exceptional scalability, availability, bandwidth and performance. To accomplish this, the SAN incorporates switch fabric technology, commonly referred to as the SANs fabric, to connect any server to any storage subsystem.

In some configurations a SAN is also connected to the network. SANs' use special switches as a mechanism to connect the devices. These switches, which look a lot like a

normal Ethernet networking switch, act as the connectivity point for SANs. Making it possible for devices to communicate with each other on a separate network. SANs are dedicated network that ensure access to application and system, and provides continuous data availability. This concept include virtualized storage, shared high-bandwidth data access, and real-time backup are now finding their way into the main-stream of distributed, open-systems computing. SANs can be implemented and collaborated in many environments such as Windows Platform like Win2K; WinNT; UNIX; Linux such as Mandrake, Red-Hat; Novell Netware; MacOS and Sun Solaris.

1.2 PROBLEM STATEMENT

Storage is an important factor in business operations. Data in business are one of the assets in the business. So, data must be kept properly and securely in order to make the business run efficiently and effectively. Information Technology (IT) plays as a key role in supporting business operation. Using an inappropriate and obsolete storage medium in IT can cause inefficient and huge drawbacks in the business itself. Costs, availability, architecture of the storage are some of the criteria in determine best storage application in the business.

1.2.1 Problem Identification

Traditionally, storage devices are attached to a server or cluster of servers via a SCSI interface. Each server attached to a network required its own dedicated storage and some new servers need to add dedicated storage for that server in order to meet the processing bandwidth requirements of a network. Besides that, if an additional storage is needed in order to support the demands through one server, additional storage would be added specifically for this server, even though an adjacent server may have excess unused storage. The software to manage storage such as tape or optical library management and hierarchical storage management are all run on the same server. Access to the storage is made exclusively through the server.

1.2.2 Significant of the Project

The expansion of IT applications has resulted in more and more servers producing more and more data, making it necessary to manage high-capacity storage systems (disks). A great deal of knowledge and expertise are needed to construct and operate such storage systems, but there are only a few people who meet these requirements. If the storage is connected in a disorderly fashion to different types of server, free space is wasted and it becomes difficult to manage. This state of affairs has given rise to the concept of integrated management, where the storage is consolidated and shared. Furthermore, increases in the volume of data make it harder to back up, and a need has arisen for systems where back-ups can be performed effectively.

1.3 OBJECTIVE AND SCOPE OF STUDY

The project aims to design best storage architecture, Storage Area Networks (SANs) which is cost effective and efficient for business operation. The project will take University of Technology PETRONAS (UTP) as the business entity since UTP has plenty of servers and clients which are running 24 hours a day a week. The study will focused on the several criteria in implementing SANs in business environment such as cost, availability, architecture, and effectiveness of usage. This new storage architecture will accelerates and safeguards enterprise processes such as online transaction processing (OLTP) applications, mission-critical backups and multimedia traffic feeding a Web site.

1.3.1 The Relevancy of the Project

Providing sufficient disk space is only part of the equation. In fact, storage is so inexpensive that IT administrators give little thought to add 50 Giga Bytes (GBs) here and there when their servers run low on disk space. But as the server farms grow, the overhead associated with directly attached storage balloons out of control, causing administrators to manage data reactively. SANs let them manage virtually all their

storage needs proactively while creating the high availability required by the server. For e-commerce and companies with extensive enterprise resource planning (ERP) implementations, server farms have become the heartbeat of operations, and missing a beat could have fatal consequences.

1.3.2 The Feasibility of the Project

The study will be divided into two areas which are research on Storage Area Networks and SANs Knowledge Management System. Research on Storage Area Networks will cover on introduction of Storage Area Networks, the advantages of SANs in business environment, hardware and software needed in implementing SANs, cost of SANs, analysis of return on investment (ROI) of SANs to business environment, SANs topology and Storage Management System. Research also will discuss on University of Technology PETRONAS (UTP) as the business entity in the research. As UTP become a big educational institution and Information Technology (IT) plays an important constraint in supporting the operational purpose, storage needs in IT has growing up. The operational purpose such as database storage, application programs, internet and intranet services, and other IT related usage. Storage in the business and educational sector in UTP has become more complex and the traditional storage technology that is being used in UTP cannot support the growing of the storage usage. In order to make UTP more flexible in term of data storage, SANs is the best solution in changing the traditional storage technology which is still being used in UTP.

SANs Knowledge Management System is a system that will support in designing best SANs solution in the preferred business environment. The system enable user to make choices in designing the best SANs implementation in their working places. The system is based on Artificial Intelligence (AI) searching algorithm and rely on Visual Basic 6 as the front-end component of the system. The expert system in the system will help the user in determine the flexibility, scalability, availability and the performance of SANs in the business environment. The expert user will give benchmark of the SANs solution to the business either it is suit to the business or not.

CHAPTER 2

LITERATURE REVIEW AND THEORY

2.1 DISASTER RECOVERY

According to Network World, [1]
Business continuance - the ability to get your network up and running after a disaster with minimal impact on business - has emerged as the leading driver behind storage-area network and network-attached storage adoption in the post-Sept. 11 world". People now realize that SANs are a key to gaining disaster tolerance," says Ron Totah, technical marketing manager for Brocade Communications Systems. Until Sept.11, network managers didn't envision the magnitude of the catastrophe that many businesses in New York faced. Totah says managers are not only planning to duplicate networks, but also are tripling them in case the original resources are destroyed.

Businesses that use a SAN to mirror data in real time over optical connections can recover much faster than those that are still sending reels of data and performing hourly synchronization processes. SANs that support longer distances offer an advantage by letting companies store data farther away.

2.2 NETWORK ATTACHED STORAGE AND STORAGE AREA NETWORKS

There has been much nonsense said and written about the differences between NAS and SAN. In fact it's probably more important to appreciate how these two technologies share more than a little in common and are in fact moving towards a common future.

According to Andy Walsky, European Business Manager, Network Storage for Quantum, [2]

There should not be any confusion between the two NAS and SAN as at present they are aimed at two fundamentally different types of user. "Quantum believes that there won't be a conflict between NAS and SAN in the future as each technology provides different, and complementary, storage functions for its users. SAN, by its nature is a more complex, high-cost solution and is relatively difficult to implement. It is aimed at enterprise level storage and is usually kept at the heart of an organization's IT infrastructure, handling 'heavy duty' storage requirements. In contrast, NAS solutions, such as Quantum's Snap family, handle workgroup storage needs and are much easier to install, you just plug and play. They take the place of adding traditional servers to the network just to increase network storage capacity."

Anthony Jones, European Marketing Manager for HP, [3]
NAS and SANs are complementary and that for the end user the most suitable choice will be based on the business benefits sought. "NAS is 'appliance' based, an optimized single function solution that provides a quick, easy way to add & manage file storage, primarily for clients, utilizing the existing network. NAS is server independent which can improve availability, and provides common storage that can be shared across different operating system environments. Jones also expresses an opinion held by many vendors of "traditional" storage technologies, saying, "On the other hand, as SAN is 'component' based, it needs to be 'built'. It utilizes a dedicated, high speed storage network, and is designed to provide 'block' level data at high speeds, primarily to application servers. It is highly flexible and scalable, but also requires sophisticated IT knowledge to implement, especially when compared with NAS."

This perceived need for sophisticated management applications has dissuaded many prospective SANs users moving from direct attached storage to networked storage because they do not want to be burdened with yet another set of unwieldy management

tools that fail to deliver. However, backup specialists such as Veritas are already on the case, providing SANs and NAS data management tools that remove such uncertainties.

2.3 SERVER CAPABILITY

According to Law firm Clark, Thomas & Winters of Austin TX, [4]

Developed a SAN in order to add storage and cluster the organization's many servers. The firm had fairly typical document management requirements. They rely on a GroupWise document management system plus email and accounting systems. "We have 300 users who need 24/7 support," says Tony Armendariz, Clark, Thomas & Winters' information systems manager. "It is important to make sure we have full fail over. The firm did not want to have to worry about a server going down. "The law firm installed a Magnitude storage area network system from Xiotech, Eden Prairie, MN. The simple, single-box system incorporates 216 GB of RAID storage and supports clustering of the firm's Novell servers. Armendariz didn't use virtual file system software because the firm wanted to maintain separate servers for each department. "The SAN improved performance and reduced the number of server crashes," says Armendariz. "Our ultimate goal is to have ten servers with five of them in a single cluster. "The firm plans to replace its current SCSI DLT tape drives with fiber channel drives that will enable continuous live backups. This will ensure 24-hour access to data with backups that are never more than a few hours old.

2.4 SERVERLESS STORAGE IDEA

According Mitt Jones, [5]

Too much data, too little downtime: For IT professionals, the challenge of efficiently backing up information gets more difficult each day, thanks to growing data repositories and to the increasingly common expectation that data be available around the clock. On top of that, it's assumed that backing up all this information shouldn't be a drag on application performance. Some systems administrators are finding relief by deploying Fiber Channel-based storage area networks, which can move live data and backup traffic off the LAN. But even with LAN-free backup, a problem persists: The server

hosting the data being backed up must move the data to the tape library across the SAN, eating up CPU bandwidth that could otherwise be used by applications.

The solution may well be serverless backup, which shifts the responsibility for data movement from application servers to some other component on a storage network, such as an intelligent router or switch. That is just beginning to happen the products such as Computer Associates' ARCserve 2000 storage-management software, for example, can be paired with hardware such as CrossRoads Systems Inc.'s 4x50 line of storage routers to send data directly to an archiving device, bypassing the application server. For companies that have already invested in a SAN, moving to serverless backup should reduce the load on application servers for a relatively small additional investment. Pricing for CA's ARCserve 2000 software, for instance, starts at about US$700. For companies still contemplating a storage network, serverless backup may provide additional impetus. "Serverless backup is the killer SAN application," says Peter Malcolm, CA's director of development for storage management.

2.5 SANs CAPABILITY

According to Hoot Thompson, Patuxent's senior technologist, [6]
SANs generally have multiple layers of hardware and software that must all work together, so most SANs are installed by specialized integrators. One such integrator is Patuxent Technology Partners of Clarksville, MD. Patuxent installs SANs for corporations and government agencies that require high bandwidth access to storage. One challenge for every integrator dealing with SAN is combining new and existing equipment, says Hoot Thompson, Patuxent's senior technologist. "We simply put a fibre channel as interface on the existing servers and connect them to the fibre network," Thompson says, explaining Patuxent's answer. "The software and file system on the legacy servers remain intact. A virtual file system like ADIC's Centravision or Tivoli's Sanergy makes it available to the SAN". Patuxent's approach lets companies leverage their existing investments in optical storage and tape archives. The basic cost for a

SANs starts at US$75,000 to US$100,000, according to Thompson. He says this cost covers a scalable framework upon which you can build an enterprise. With SANs scalability, you need only start with enough storage for your immediate needs. Storage gets cheaper every day, so it makes sense to put off purchases until they're actually needed.

2.6 SANs ADVANTAGES

According to Skip Jones, president of the Fibre Channel Industry Association, [7]
"SANs let you to pool the storage so that multiple servers can reach data, easing scaling and reducing management costs," says Skip Jones, president of the Fibre Channel Industry Association. Jones says that SANs operate independently of the local and wide area networks that users connect to. This enables databases and other applications to access data without having to compete with user traffic for bandwidth. Also, by putting all the storage in one place, an administrator can manage it from a single console rather than logging into multiple independent storage servers. Scaling can be as easy as plugging a storage device into a network port.

According to Eric Herzog, vice president of marketing at Mylex, [8]
Despite the advantages of a SAN, don't go charging into building one without careful thought. "SAN components are expensive," warns Eric Herzog, "Fibre channel host adapters are two to three times as expensive as the SCSI adapters; cabling, hubs and routers are about twice as expensive as components used in traditional networks."

CHAPTER 3

METHODOLOGY / PROJECT WORK

3.1 METHODOLOGY

Research and developing the SANs system need student to adapt software engineering paradigm as a discipline that integrates the process, methods and tools in the system development. Student used System Development Life Cycle (SDLC) to ensures consistency and reproducible in the development area. SDLC also reduce risk associates with mistakes and shortcuts and enable to produce complete and consistent documentation for the projects. Planning, Analysis, Design and Implementation are four basic terms in SDLC. These terms are used according to SDLC model such as Waterfall Model, Spiral Model, Hybrid Model, or Prototyping. Each model has own advantages based on the project specification and requirement.

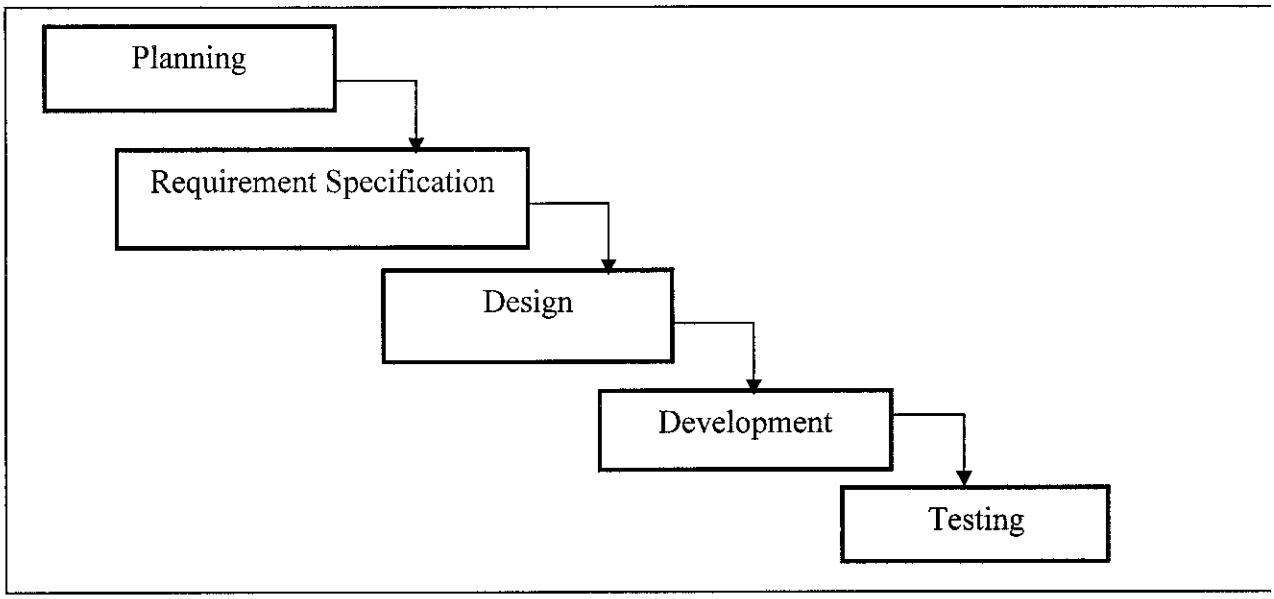


Figure 1.0 : System Development Life Cycle Model

Building a system based on the network application needs a repetitive model combined with prototyping. Hybrid model is the best project development life cycle methodology. Iterative development ensures system is developed according to the module and constant review and testing are key element in the development process. Therefore hybrid model is suitable in the development process which allows the student to review each stage in the development process and testing procedures upon to the network environment model.

The development process begins with the student playing the role of the knowledge engineer in defining problems, objective and requirement. This is achieved by soliciting the domain expert on the knowledge where knowledge plays an important part on the development. Knowledge is both the understanding on the main problem and the rules to solve it. Network procedures and student's supervisor is the domain expert that will serve as the main reference in understanding the requirement and procedure.

Design and module development are the next procedures in the development process. Knowledge engineer will represent the knowledge acquired in computer and prototype is developed to solve problems according to the network environment. Knowledge engineer and the domain expert will review each completion of the modules and integration part will be conducted as the modules are completed based on the requirement.

3.2 HARDWARE AND SOFTWARE REQUIREMENT

Some factors in choosing hardware and software to be used in designing and developing the system are:

3.2.1 Operating System / Platform

Operating system is the major impact in designing and developing the system where the system must be supported with the current operating system which being used in the

organization. MacOS, LINUX, UNIX, Solaris and Windows platforms are the major operating system in this world. For this project, the student has taken Windows as the Operating System platform.

3.2.2 Executable File Programs

The SANs KMS is an executable file where the OS must support .exe programs file. SANs KMS will come out with setup preferences in order to install to the computer.

3.2.3 Database Support

Microsoft Access is chosen as the database platform since the compatibility of the system with the database platform.

3.2.4 Technical Skills

The technical skills in using the system have to be identified earlier. Normal users, expert users and administrators are the users in the SANs KMS and their technical skill would be different from each other. It is important to consider the level of easiness of learning and familiarizing to the system.

3.2.5 User Ownership

The user has to identify type of the user such as the system administrators, expert users and the normal users. The preference for the system is customized based on the user's type. Further more, there are different of data ownership in the system.

CHAPTER 4

RESULT AND DISCUSSSION

4.1 RESULT AND DISCUSSION

The project aim to deliver a system in SANs which will give the best SANs architecture for the business. Besides that, the system also will able to do the SANs configuration and debugging through the web application. Collaboration and artificial intelligence (AI) are the main aspect in the SANs system. The system will deliver a best solution in SANs architecture with the collaborative element to support configuration and debugging procedures in the SANs. AI plays the knowledge management system in the SANs. Below is the identified system requirement for product development and implementation.

Specification	**Type**	**Version**	**Licensing**	**Developer**
Testing and Implementation Server	Windows Platform	XP	Standalone	Microsoft
Artificial Intelligence Knowledge Based Engine	SWI PROLOG/ LPA-Win Prolog	5.2.6 / 4.320	Standalone	SWI Prolog / LPA-Win Prolog
Relational Database	Microsoft Access	XP	Standalone	Microsoft

Table 1.0 : System Development Requirement

4.2 STORAGE AREA NETWORKS BENEFITS AND FEATURES

4.2.1 Storage Devices Capabilities

One of the major advantages of a SAN is that it brings all the storage into one place. While the server still has transparent access to the storage devices, all the disks can be centrally managed from one location. By putting all the storage in one place, an administrator can manage it from a single console rather than logging into multiple independent storage servers. The amount of disk space mapped to each server can be increased, moved to a different RAID striping scheme or reassigned to other servers' on-the-fly, without bringing down the SAN By sharing storage pools in a SAN, servers sharing data sets can fail over seamlessly. Replacing traditional SCSI DLT tape drives with fibre channel drives will enable continuous live backups where it will ensure 24-hour access to data with backups that are never more than a few hours old. The high-speed Fibre Channel infrastructure between storage pools also can reduce disaster recovery from several hours to less than a few minutes. When coupled with an automatic fail over using multiple servers, HBAs, SAN switches and disk arrays, an enterprise can create a very robust and resilient storage system. The tape libraries connected to the SAN can back up and restore the disk arrays using the Fibre Channel network without affecting the servers or their applications.

4.2.2 Server Advantages

The SAN improved performance and reduced the number of server crashes. Business owners did not have to worry about a server going down. Nowadays, businesses have move from using virtual file system software to the server clustering. The purpose is to maintain separate servers for each department, for example to have ten servers with five of them in a single cluster. This will increased efficiency of data utilization due to the sharing of data between different types of server. Dedicated storage paths in the network ensure greater system flexibility and expandability due to the freedom of access between all types of server and storage devices.

4.2.3 SANs Connection

SANs employ some of the most advanced features of fibre channel technology. Fibre Channel technology has ability to operate over optical transmission lines for long distances using fiber optic channels and can be used for the storage connection. Since the connections between the servers and the SANs are through fiber-optic links, the various parts of the system can be located hundreds or thousands of feet apart which is useful for storage consolidation. Fiber Channel also enables databases and other applications to access data without having to compete with user traffic for bandwidth. SANs are useful for corporations and government agencies which require high bandwidth access to the storage. Fast network connections improved storage expandability, device flexibility and data access speed.

4.3 STORAGE AREA NETWORK ARCHITECTURE

4.3.1 Basic SANs Implementation

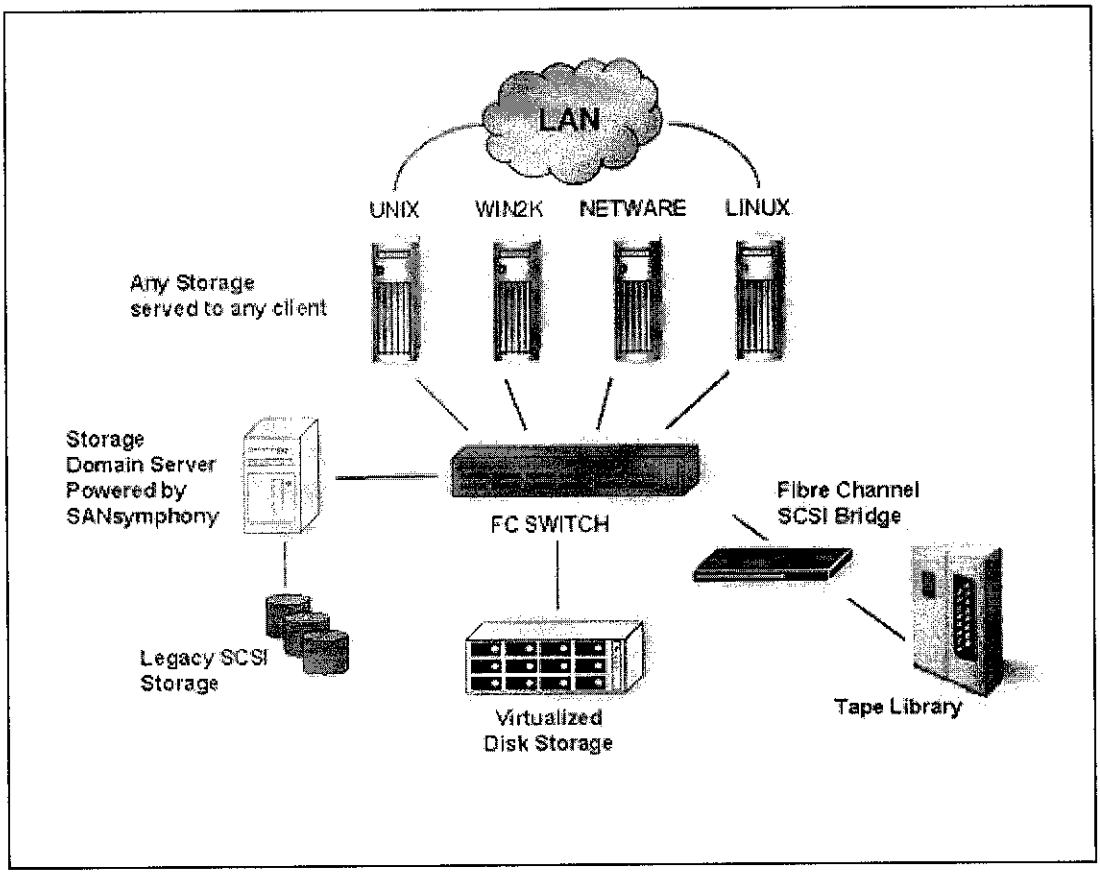


Figure 2.0 : Basic SANs Architecture

For environments of 500GB or more, a virtual storage environment should be strongly considered. Mechanisms for protecting existing investments in SCSI storage are included. Starting with an economical switched Fibre Channel system, a low-cost approach to a SAN can be implemented. Whether to move to a Fibre Channel (FC) based SAN environment while protecting the existing investments in legacy SCSI based storage or to get started on a new infrastructure within a limited budget, the above model can provide the basics.

The hardware infrastructure incorporates suitable Host Bus Adapter (HBA) cards in each server. These can incorporate either copper of fiber optic connections, depending on the distances involved. Generally within a data center, connections up to 20 meters can be accommodated on copper.

The core of the infrastructure is a switch that provides the connections between the client application servers or workstations and the SAN. With environments above about 500GB, a Storage Domain Server (SDS) implemented on open systems hardware can be used to virtualize the storage, so that a mixed OS environment or group of individual storage clients can be allocated storage as needed from a common pool. Using a virtual storage architecture can improve utilization by at least 30%, while the large cache on the SDS markedly improves transaction performance.

The FC/SCSI bridge also provides a connection to SCSI-base taped libraries for back-up. Using the optional snapshot capabilities of the SDS can allow real-time backup using conventional back-up packages.

4.3.2 Workgroup SANs Solution

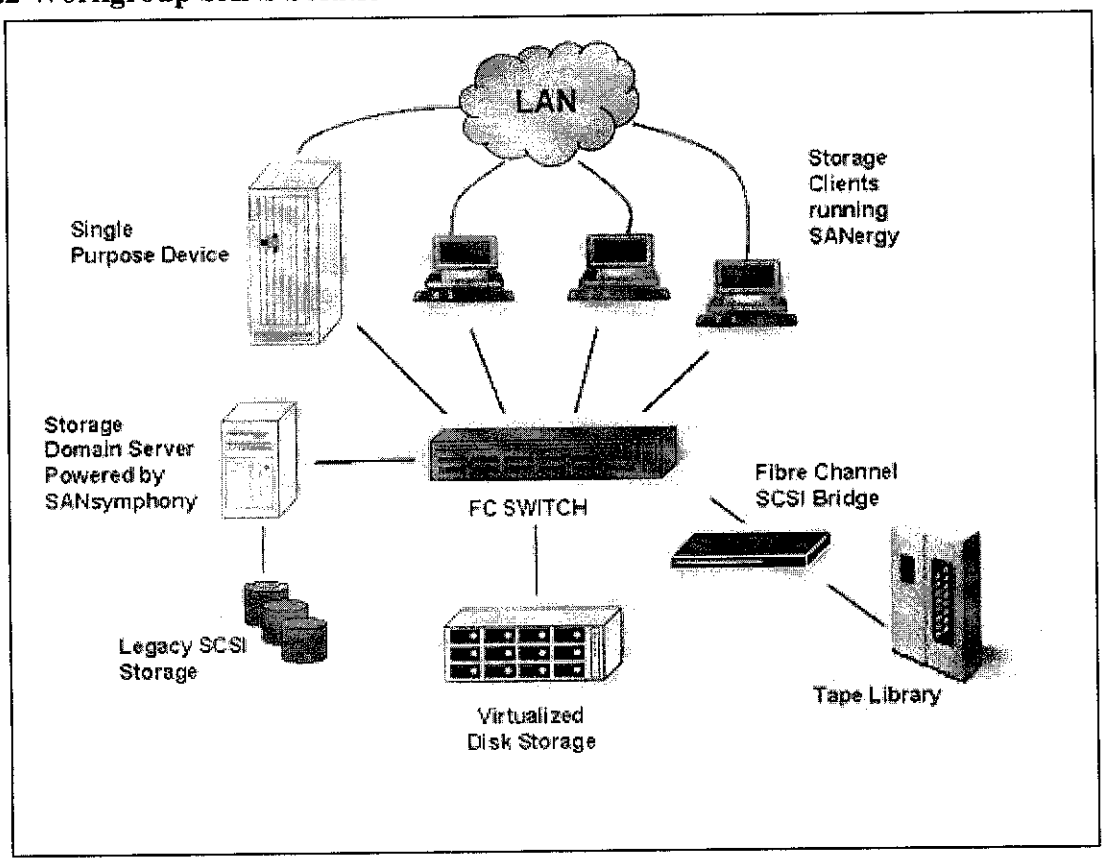


Figure 3.0 : Workgroup SANs Architecture

A major benefit that SAN technology provides is the ability of a group of workstations to connect directly to a shared storage resource, without the use of servers. The SAN storage can be virtualized, or simply zoned through the switch in a small installation. Data flows directly to the workstations on a Fibre Channel link, without LAN or server bottlenecks. Pre-press, Video Editing, and Engineering group applications can benefit from this approach.

The above Workgroup SAN Solution represents a template for allowing workstations in a group have access to a shared data set without a dedicated server. This scenario can be implemented on a basic SAN involving only a few workstations with or without Virtualized Storage. It can also be extended to a very large installation with 20, 30 or more workstations sharing access to the same data on the SAN whose storage is

presented from a common pool using Storage Domain Servers (SDS's) running DataCore's SAN symphony on open systems hardware.
A workgroup may be involved in electronic pre-press operations, video editing, or engineering design. Shared access to the data is provided through the use of Tivoli SANergy redirector software that runs cooperatively on the attached workstations. Only one of the workstations need be designated as the meta-data controller; however, failover options can be defined, or a designated (WinNT, UNIX, LINUX) device can be used as the meta-data controller.

The workstations in the group need not be of a homogeneous type. A mix of WinNT, MAC and UNIX/IRIX/LINUX workstations can co-exist and share a common data pool. No more "sneaker-net" for file transfers. In fact, no more file transfers are required in the conventional sense at all, because the data sets are logically available to each concurrently as if they were locally present.

In some environments a special purpose device may also share the SAN connections and use the output of the workgroup as its source of data. For example, in a pre-press operation, the single purpose device could be the Raster Image Processor. In a video production group, the device may be a digital video recorder or server. The physical SAN infrastructure requires Host Bus Adapter (HBA) Cards in each workstation, and connections to a Fibre-Channel storage switch.

Optimally, the storage should also be composed of Fibre Channel disk drives; however, a bridge to legacy SCSI storage is possible. For data redundancy, data transfer bandwidth enhancement, the drive enclosures should incorporate a Fibre to Fibre RAID controller. InfraStor offers disk enclosures with RAID controller options that are capable of up to 190 Mbytes/sec reads and 150 Mbytes per second writes. Their scalability is unsurpassed. Start with a few drives in a JBOD arrangement, and as needs increase, cascade successive 12-drive enclosures for multi-terabyte storage capabilities. Note that legacy SCSI storage can be virtualized as well.

Finally, if the data is valuable enough to designate a special work-group to create it and manipulate it, there should be a means to back it up. A tape library of sufficient capacity, couples with the optional Snapshot capabilities of SANsymphony can be used with many well-known back-up solutions.

4.3.3 Enterprise SANs Solution

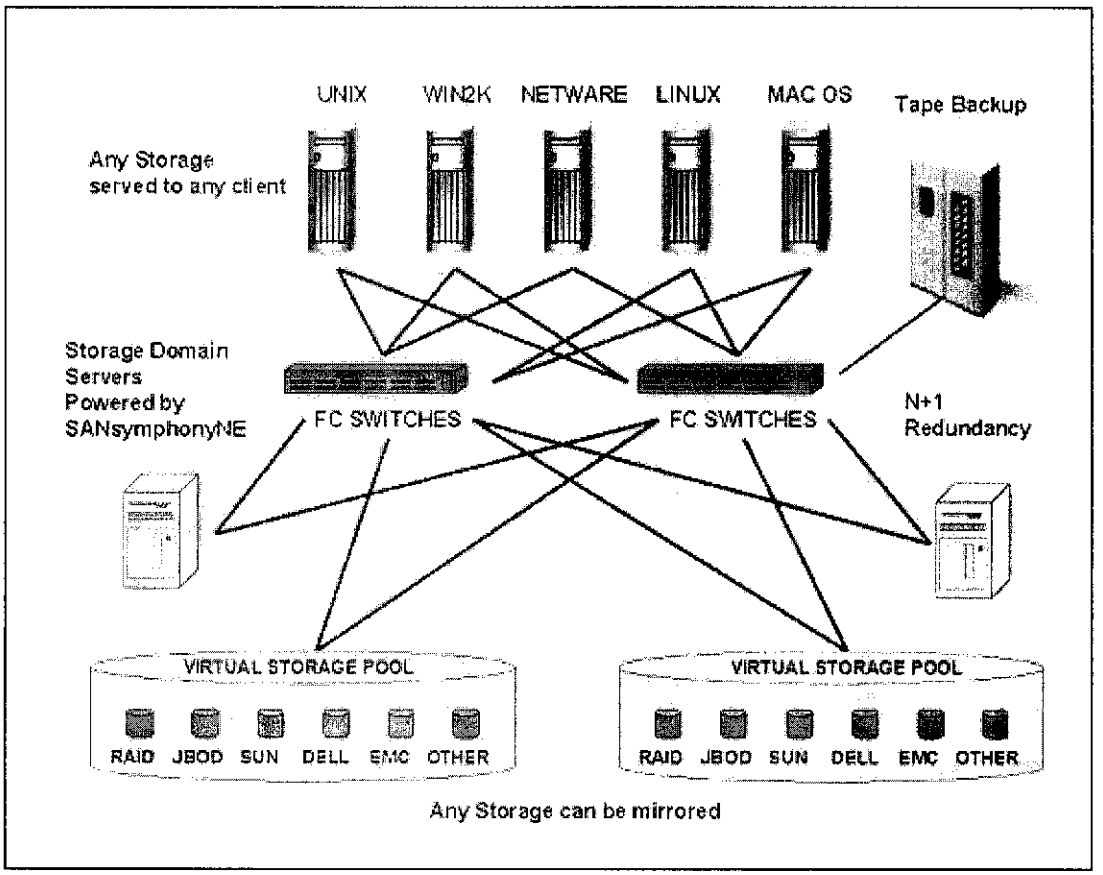


Figure 4.0 : Enterprise SANs Architecture

Scalability and high-availability are the key requirements. Virtualized Storage is the only logical approach to managing a large corporate environment. In environments where a hardware failure will lead to the shut-down of a business-critical application, the lost business costs can mount up to millions of dollars an hour. By laying out SAN architecture with redundant links and hardware elements, the effects of single point failures can be avoided. By taking advantage of the N+1 scalability of Storage Domain

Controllers the Virtual Storage environment allows for redundant data access paths to be defined.

The enterprise SAN Solution above represents a template for providing full redundancy in a Virtualized SAN Storage environment. On a connection level this is accomplished by having dual Host Adapter Cards in each server, and separate connections to an array of Fibre-Channel storage switches as a meshed fabric. Multiple levels of redundancy are available in the storage, from RAID on each enclosure to full mirroring by the Storage Domain Servers running the Network Edition of SAN symphony from DataCore Software.
SAN symphony Network Edition provides N+1 redundancy by maintaining a heartbeat between devices, and mirroring storage between storage domains. Loss of access to storage on a designated primary storage domain can be accommodated by access to the secondary domain.

All storage that can be seen by the SDS can be virtualized. This means that even existing enterprise class storage arrays can be treated as a virtualized device, thus saving on service and maintenance charges.

4.4 REASON OF IMPLEMENTING SANs

4.4.1 Scalability

Add servers and storage independently. Servers and their associated storage are no longer tightly linked together. New storage devices can be dynamically added to the central pool of storage, without having to add a new server. In a similar way, additional servers can be dynamically added, should require additional processing power for the applications.

4.4.2 Availability

Facilitates shared on-line spares and remote back-up or mirroring and reduces down-time requirements. The Storage Area Network dramatically improves availability of

business processing. For example, any server can take over from a failed applications server as they share access to the same storage and users. There are no longer constrained by bottlenecks in terms of how to get to the data and now can reach to the corporate data through any application server because of the improved clustering solutions. In addition, data can be automatically replicated to where it is needed in business terms. Technology is now becoming available that will enable such replication to occur at either the disk/volume level or at the database/file-system level. For the really critical applications a change of data value can be updated on all replicas simultaneously or synchronous replication. For most data, however, it is perfectly acceptable that replicas are a few seconds or minutes out of date to their master copy, asynchronous replication. A fundamental utility within the Storage Area Network is a set of replication facilities to meet these needs. These replication facilities must be completely accurate, reliable and exhibit very high performance and thereby give the company confidence to remove all the 'private' copies that probably exist to date. Hence the Storage Area Network model provides inherent availability of data by the use of automatic data redundancy, automatic backups and the maintenance of nearby disaster recovery copies. User-level replication, as mentioned above, also adds further resilience if provided. Clustered servers with shared access to the data that can dynamically switch users and applications between peers will dramatically improve user and business application availability.

4.4.3 Flexibility

Reconfigure Storage and Servers on the fly. Isolated data sources can be interconnected and made generally available to multiple servers. The limit for how far apart these sources can be is defined by the enabling infrastructure of the fibre channel itself. This facilitates the re-use of storage in a more efficient manner than ever before. Indeed, through the SAN architecture, consider the allocating storage from one application to another, even though the applications are running in different physical locations. The result is a far better return on investment from the existing infrastructure.

4.4.4 Performance

High performance access to global data is achieved because of the inherent performance characteristics of the enabling fibre channel technology. SAN also provides multiple servers to storage paths, for example HP-UX, HP OpenVMS, HP Tru64 Unix, Microsoft Windows NT, Windows 2000, Datacenter, SUN Solaris, IBM AIX, Novell NetWare, Redhat Linux, SUSE Linux, Caldera Linux. Besides that, the SAN architecture makes it possible to relocate various activities such as backup, restore, file migration and replication of data. Instead of passing data across the local/wide-area networks via servers, it has the opportunity to move the data directly from disk/tapes to other disks/tapes across the SAN fiber

4.4.5 Return on Investment of SAN

Scalability, flexibility, availability and performance are the main factors that provide quantifiable business benefits relating to increased return on investment (ROI). The distributed set of storage devices can be viewed a as though they were one single homogeneous mass. Hence, if a given application requires additional disk capacity, then it can re-assign disk capacity from any other application on the SAN even though these other resources may be located many miles away from the primary application. This gives the enormous improvement in ROI because the re-use storage devices that previously were hidden behind a given server.Besides that, the more distributed the environment, the more costly the maintenance in terms of staff resources. The bottom line shown in the chart to the left is that consolidation of distributed NT-based storage to a virtualized SAN-based resource can save 80% or more of the costs of management.

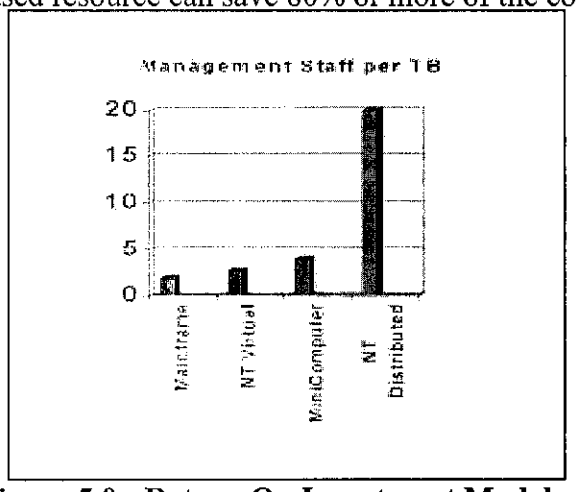


Figure 5.0 : Return On Investment Model

4.5 UTP RESIDENTIAL COLLEGE INTERNET PROBLEM

4.5.1 Problem of Internet Connection

Currently in UTP residential colleges, there are two villages that are support intranet and internet connection which are Village 3 and Village 4. Novell Netware Border Manager is the proxy servers that are being used in UTP for internet connection. For UTP residential colleges, two servers are used to support the student accommodation for internet in their rooms. But both of these villages are using separate proxy servers for the internet connection. For village 3, the proxy server is utp298bm1 which the IP proxy is 160.0.226.207 while for village 4, the proxy server is utp298bm2 which the IP is 160.0.226.208. Figure 6.0 Show the layout for the internet connection for village 3 and 4.

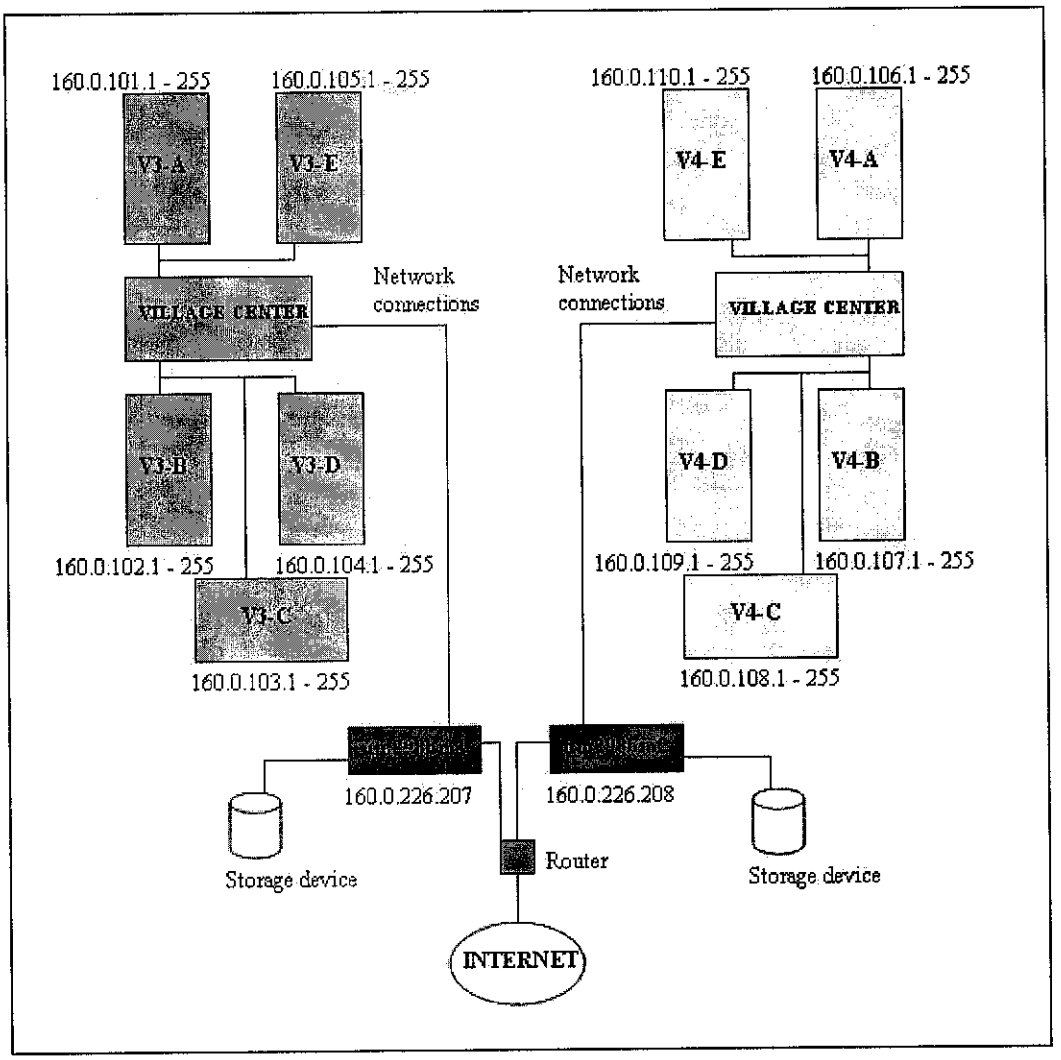


Figure 6.0 : Current Proxy Server Connection in Residential Colleges

Novell Netware Border Manager works as proxy server which connects UTP intranet with internet using router. The current architecture use two servers which are independent with each other. Each server has their own storage devices and internal backup devices. The problem occurs when the storage cannot support more cache logs and database logs. When the server cannot support more space for the logs files, the connection between the proxy servers with the internet will be terminated. The problem become more complicated when both of the server working independently from each other. The server cannot take over from the other failure server. This sometimes makes the students fail to connect to the internet.

4.5.2 SANs Solutions

Storage Area Networks is the best solution to solve the current problem in UTP residential colleges. Implementing SANs in UTP environment will cut several problems occurred in UTP. Since UTP use many servers for their business operation such as Academic servers, DNS server, email server and many more, the usage of these server become more competitive and growth due to the growing of the UTP facilities and business operation. UTP now in a process of building their intranet and internet connections since their building facilities are still in under construction. New Chancellor Complex which consists of main hall and library, and the new residential villages are still under construction. These building need internet and intranet connections and new servers must be added on in order to build new networks on UTP. Designing independent networks connection between each server may decrease efficiency of LAN capabilities. Servers and the storage devices must be interrelated with each other in order to make full usage of their capabilities and technologies. Figure 7.0 shows the proposed SANs solution and architecture for the UTP environment.

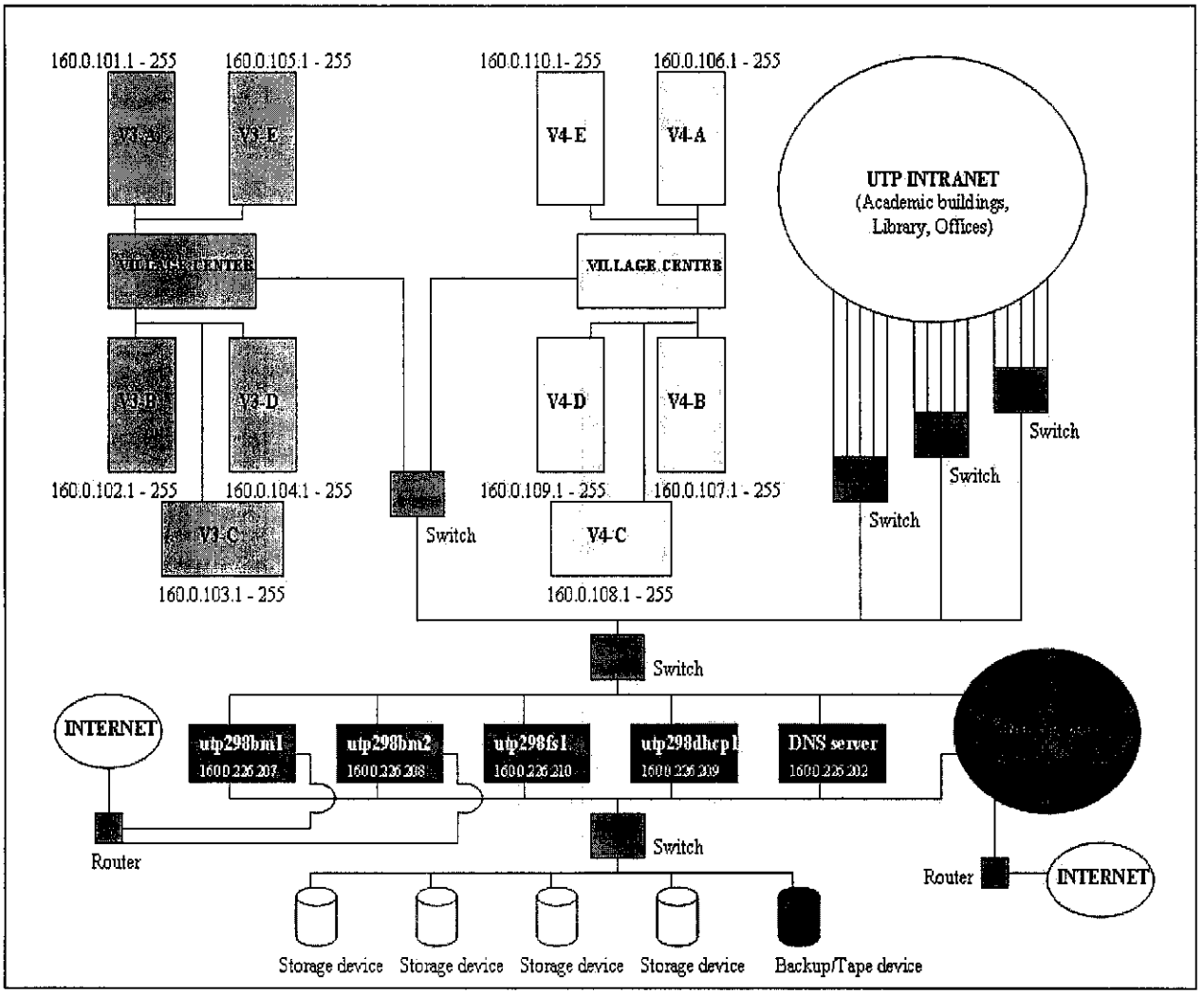


Figure 7.0 : Propose SANs Architecture in UTP

From the figure, it shows that the servers are interconnected with each other. The storage devices also interconnected with the servers. New backup device is added in order to make backup for the servers and the storage. Implementing SANs in UTP not only solve the internet problem in residential colleges, but it also makes easiness for the new server or new storage devices to be added on to the networks. For the internet problem in the residential colleges, when one of the proxy server is down, for example utp298bm2, the other server will be automatically take over the fail server. This is the mean of availability in Storage Area Networks where any server can take over from a failed applications server as they share access to the same storage and users.

Besides that, implementing SANs in UTP will make the networks more scalable. Scalability in SANs means that servers and storages are independently. The servers and their associated storage are no longer tightly linked together but still interrelated. The new storage devices can be dynamically added to the central pool of storage, without having to add a new server. In a similar way, additional servers can be dynamically added, should require additional processing power for the applications. Since UTP will require storage devices and servers for their new academic and residential buildings, adding new servers and storages become easier and flexible.

In term of ROI, when UTP implements SANs in their environment, the needs to buy new server for new buildings will drop. This will decrease the cost of buying servers instead buying or adding new storage devices for future usage. For UTP, the best SANs architecture would be enterprise architecture. This is because the growth usage for storage in their environment such as for academic databases, student databases and other application databases. Moreover, UTP is and education institution where the needs for IT facilities and internet connection is highly needed. Furthermore, inefficient in manage storages and IT facilities would destruct business operation.

4.6 STORAGE AREA NETWORKS KNOWLEDGE MANAGEMENT SYSTEM

Storage Area Networks Knowledge Management System or SANs KMS is a system that gives decision support in choosing best Storage Area Networks solution for business environment based on the cost estimation. The system consist of two part, the first part is the decision support in selecting Storage Area Networks solution for the business environment, and the second part is the expert system which gives expert review of the chosen SANs solution of the business.

For this SANs KMS prototype, the users are the SANs vendors and distributors. The users for the SANs KMS are divided into three categories which are normal users, expert users and the administrators. Each user has certain privileges. For normal users,

they don't have privileges to access the administration part and the expert system. The expert users have privileges for the expert system but don't have for the administration part. Only the administrators have the privileges to access all part in the SANs KMS.

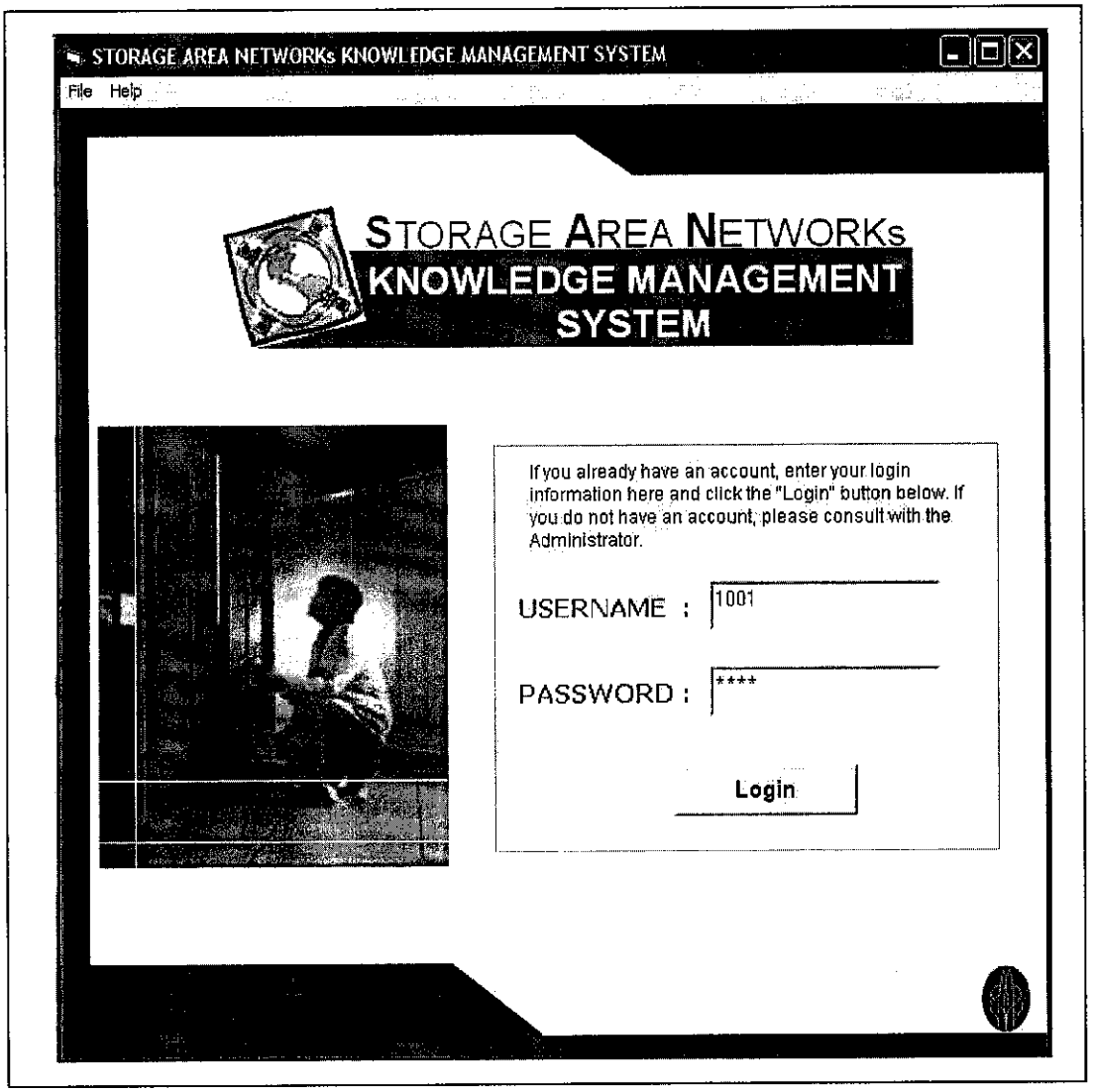


Figure 8.0 : Login Interface

Login is the first user interface. Password is required to limit access privileges.

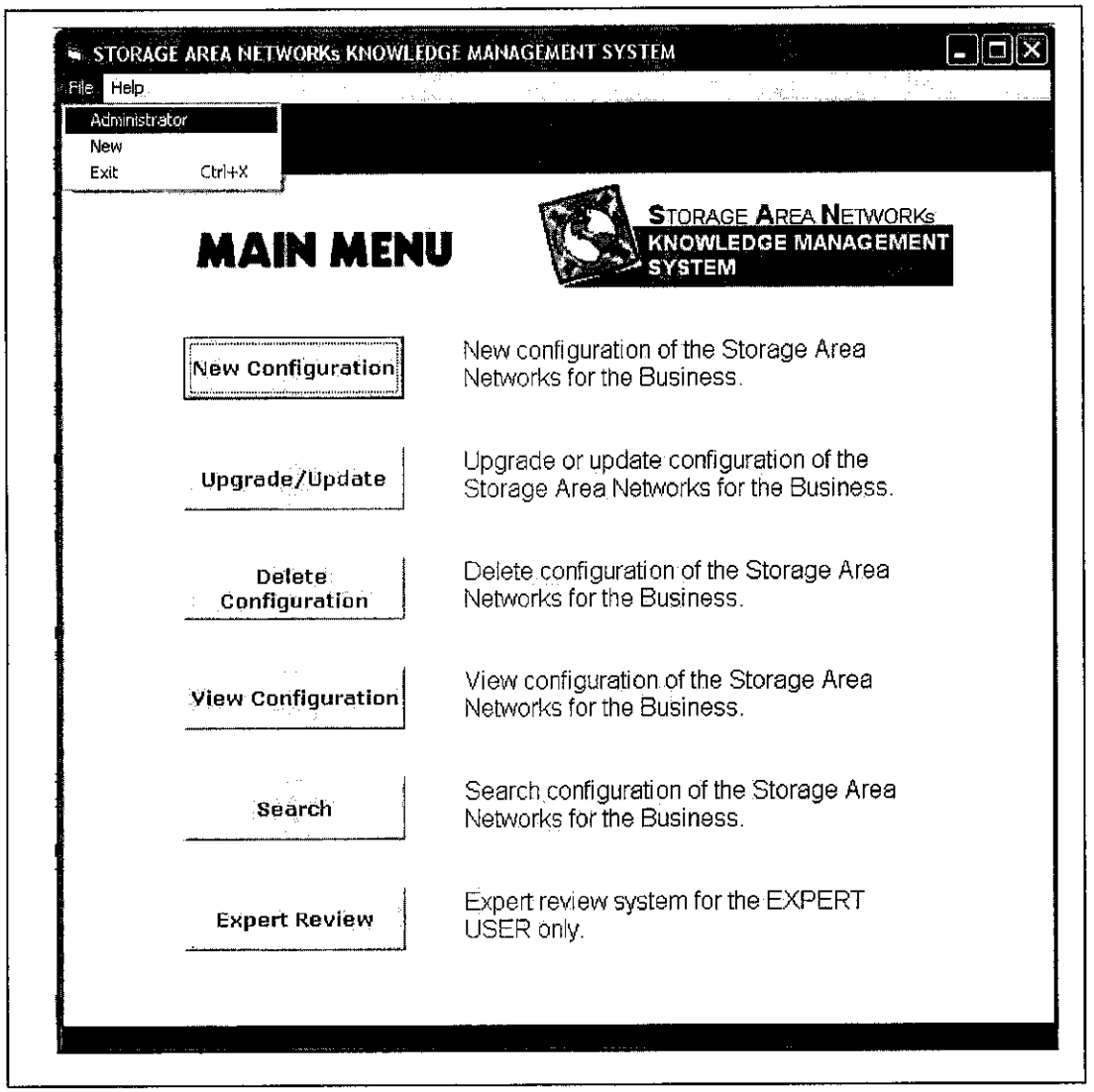


Figure 9.0 : SANs KMS Main Menu

Main menu is the second user interface in the SANs KMS. In the main menu, the users can view certain function based on their user type either as normal user, expert user or as administrator. For normal user, they don't have access for the administrator and expert review function and for the expert user; they don't have access for the administrator function. Only administrators have the access for all function in the system.

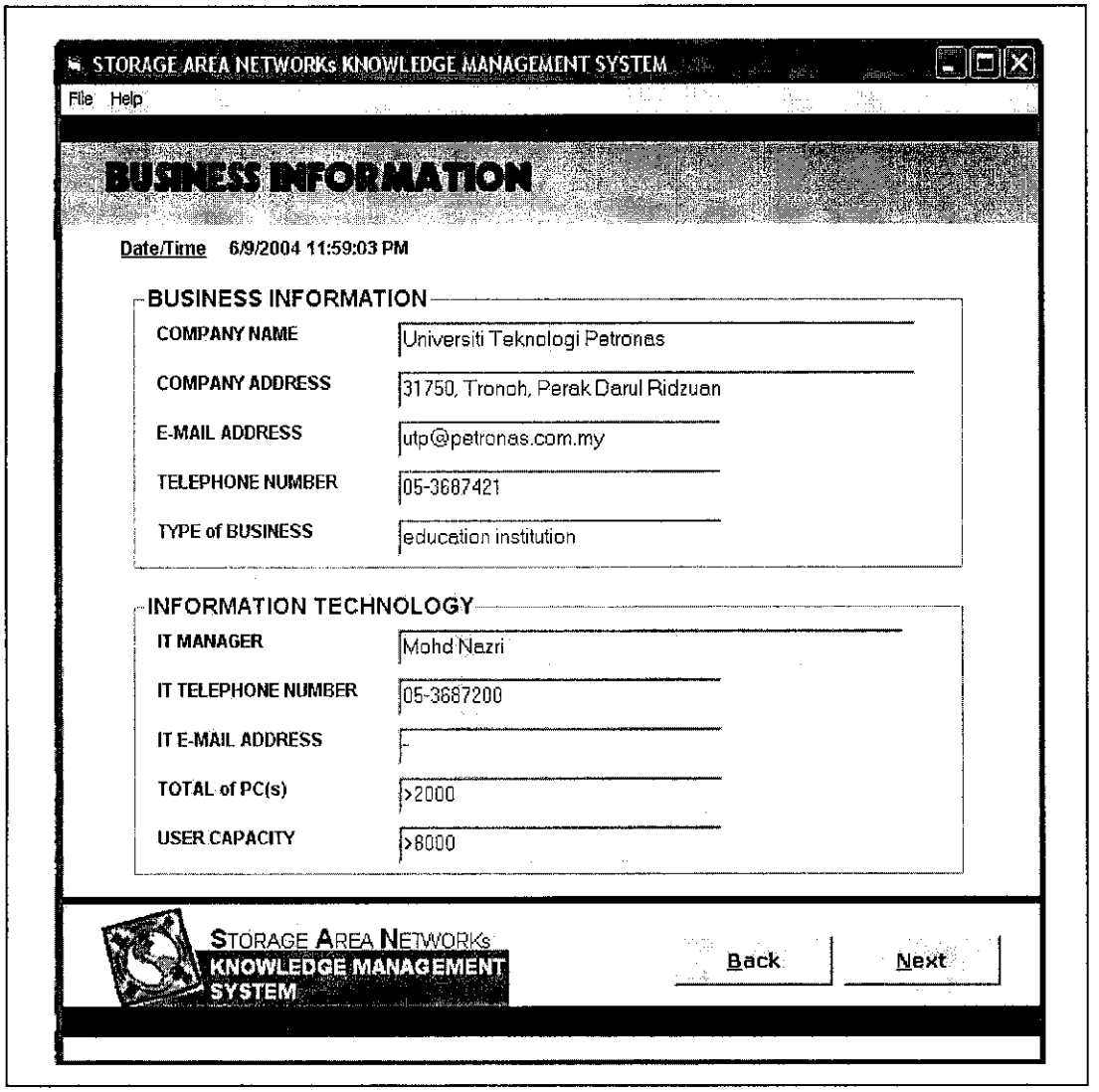


Figure 10.0 : Business Information Page

Business Information page capture the information about the business such as company name, the contact person for the IT, the user capacity of the business and many more. Besides that, the software used in the business operation such as server operating system and backup software are also capture to the database.

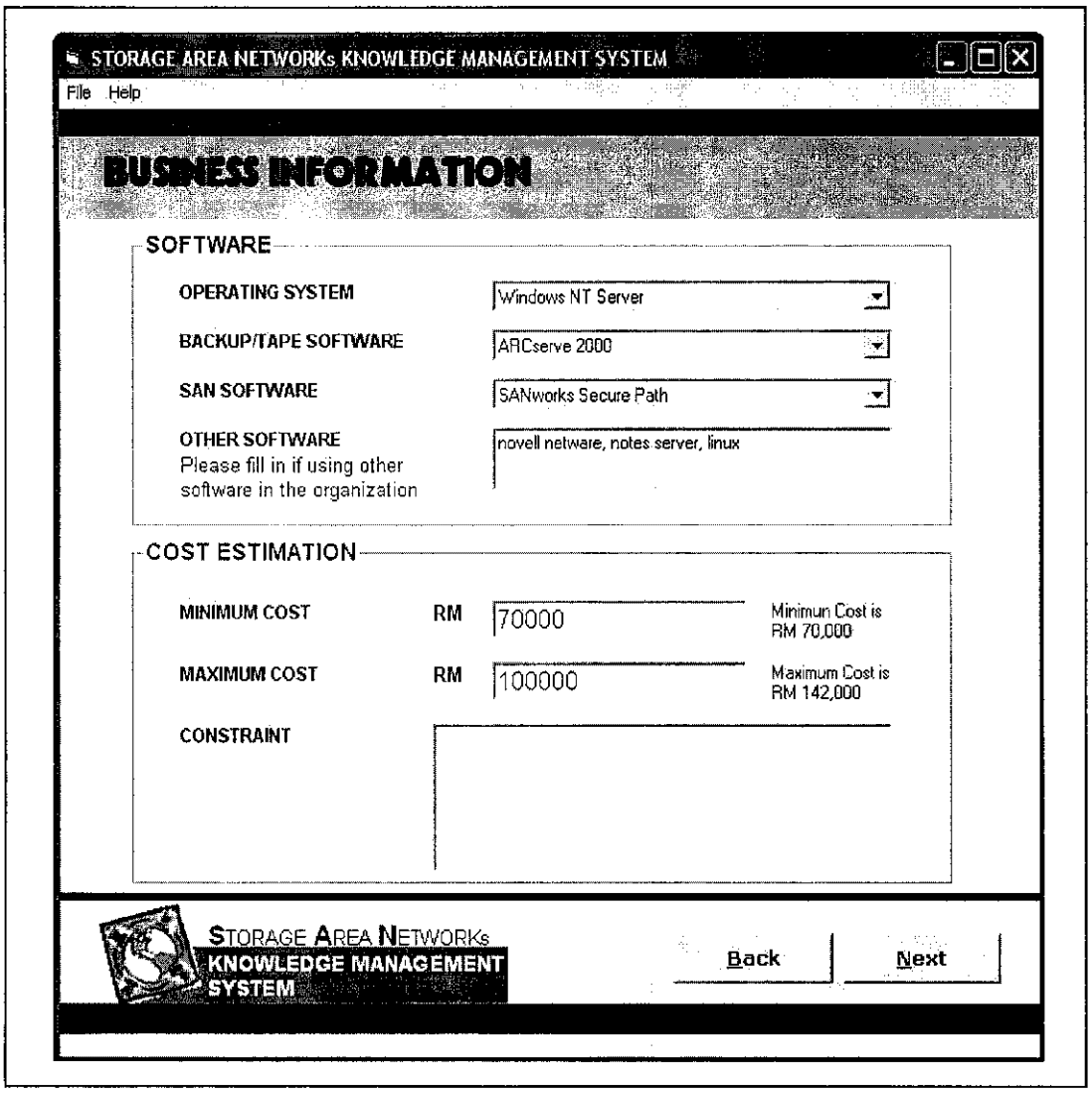


Figure 11.0 : Cost Estimation Page

The SANs solution will be based on the cost estimation which the business has budget for their SANs. Minimum cost for SANs is RM70,000 and the maximum cost is RM142,000. This cost based on the several hardware in SANs such as server, RAID, HBA and the others. From the cost estimation, several SANs solution will come out as the total price for the SANs. The price will fluctuate based on the current hardware price in the market.

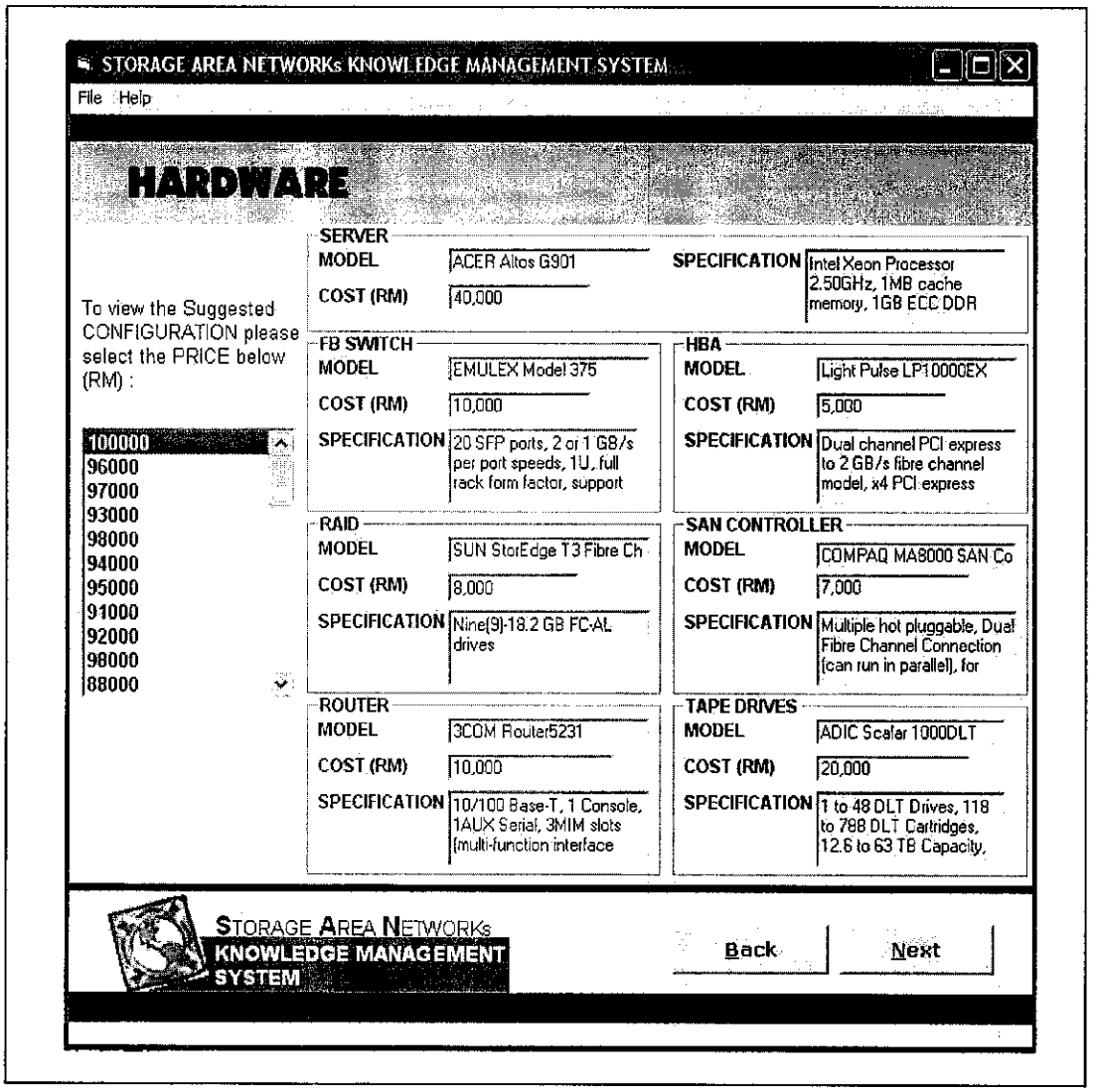


Figure 12.0 : SANs Hardware Page

The SANs solution will come out based on the cost estimation earlier. The user has to pick one of the prices which come out from the range of cost estimation. An Artificial Intelligence is used in making combination of the SANs hardware with the cost range. Each SANs price will come out with different combination of hardware.

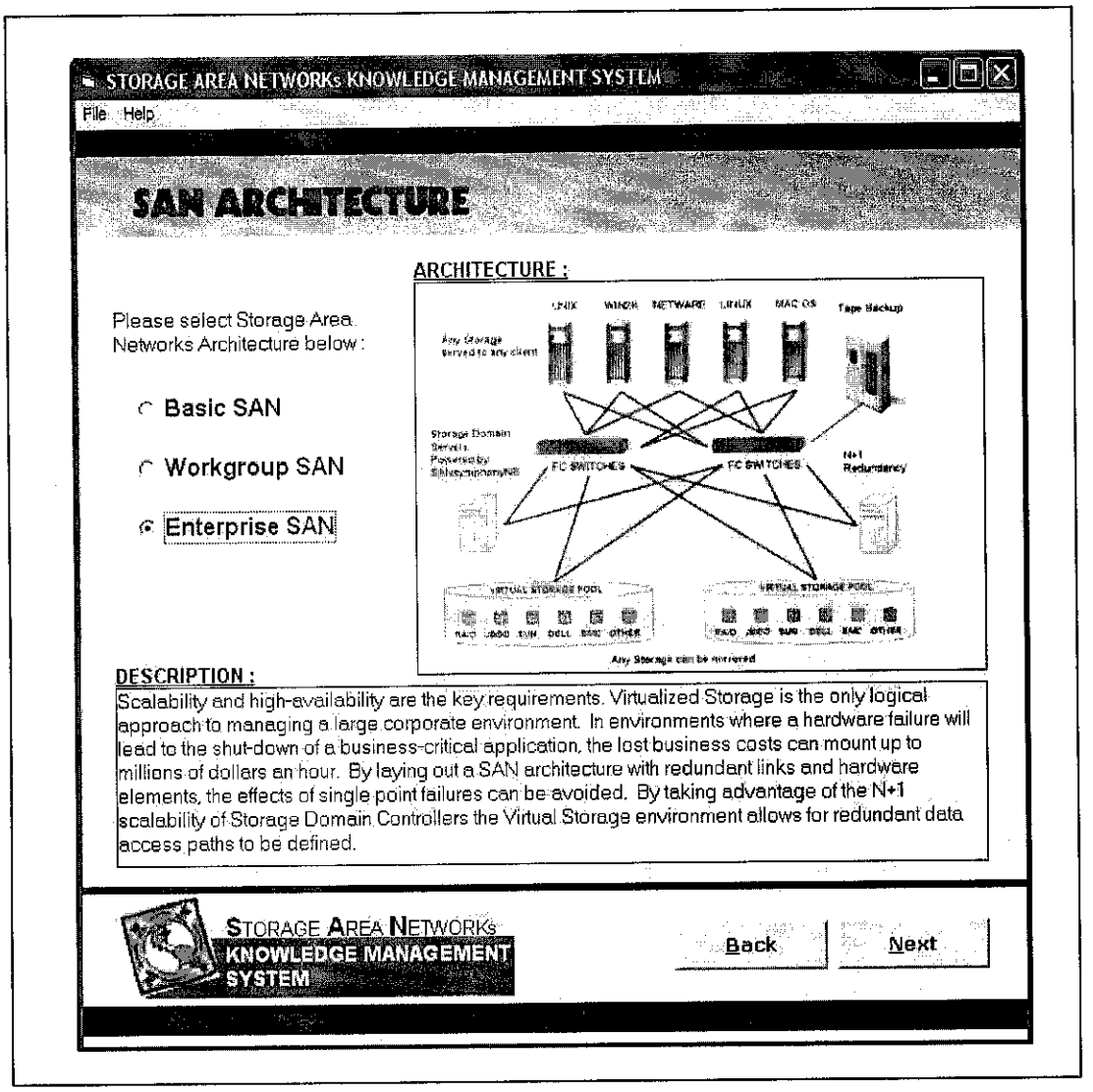


Figure 13.0 : SANs Architecture Page

The are three SANs architecture in the SANs KMS which are basic, workgroup and enterprise architecture. Each architecture has advantages and purpose. For UTP, enterprise architecture is the best solution since UTP is one of the big educational institutions and the use of storage and IT is high.

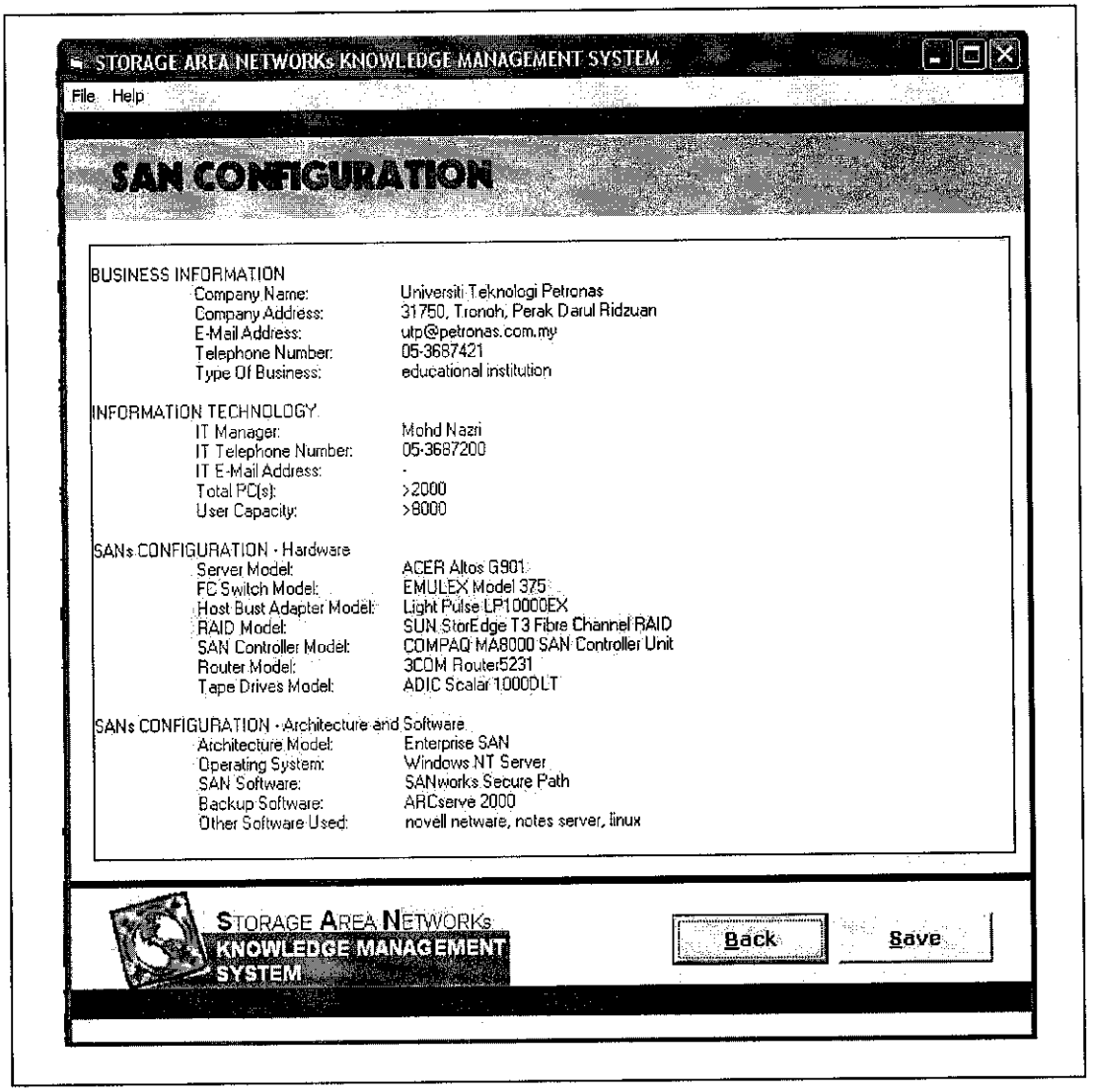


Figure 14.0 : SANs Configuration Page

SANs configuration page display the information and the selected SANs solution for the business before being saved to the database. This information will be used by the expert users in making benchmark of the SANs solution to the business based on the four criteria which are scalability, availability, flexibility and performance.

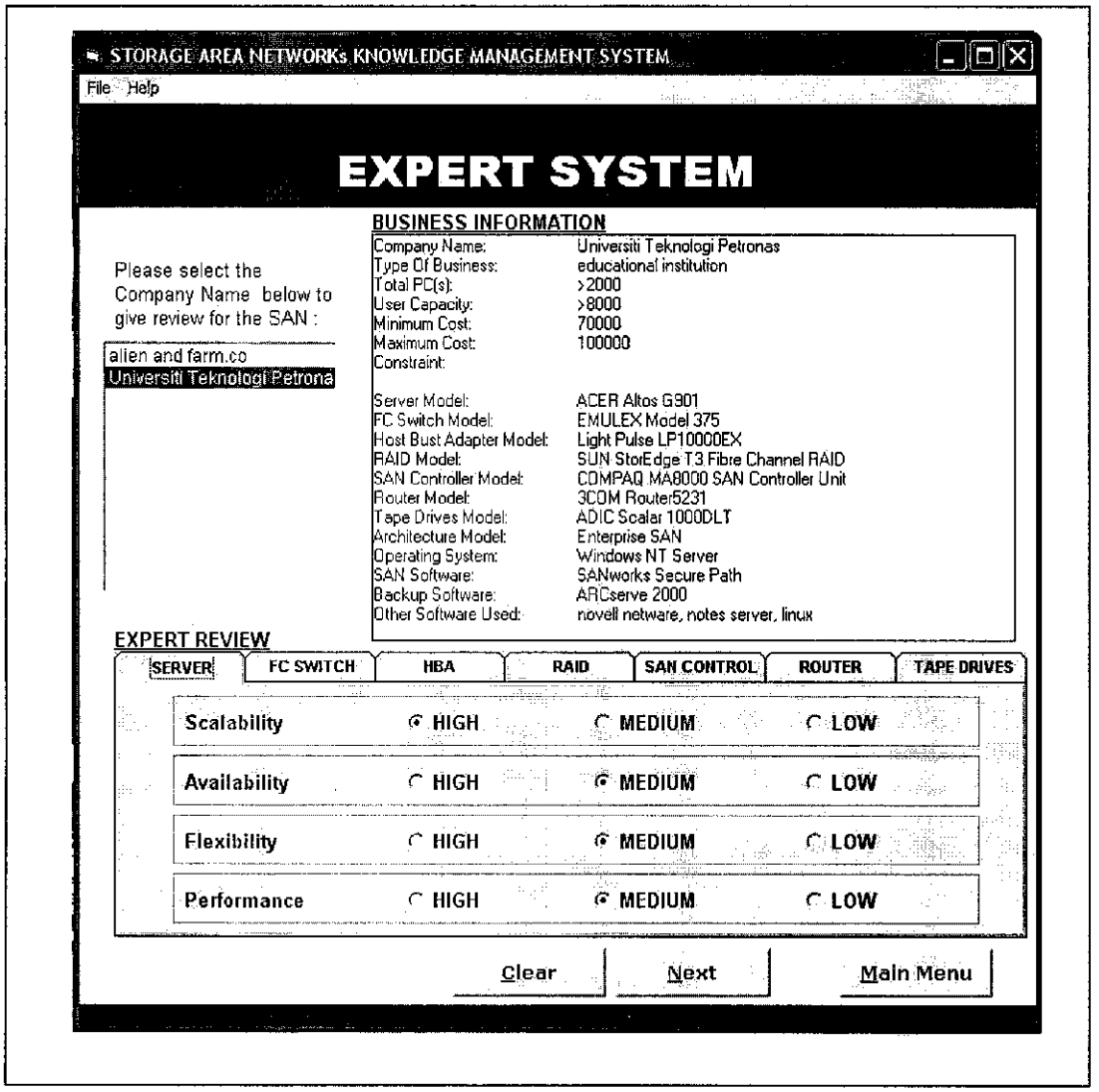


Figure 15.0 : Expert System Configuration Page

The expert user will choose the SANs configuration of the business and do the benchmark based on the scalability, availability, flexibility and performance of all the chosen hardware. High, medium and low are the three benchmark in determine whether the SANs configuration is highly suitable to the business.

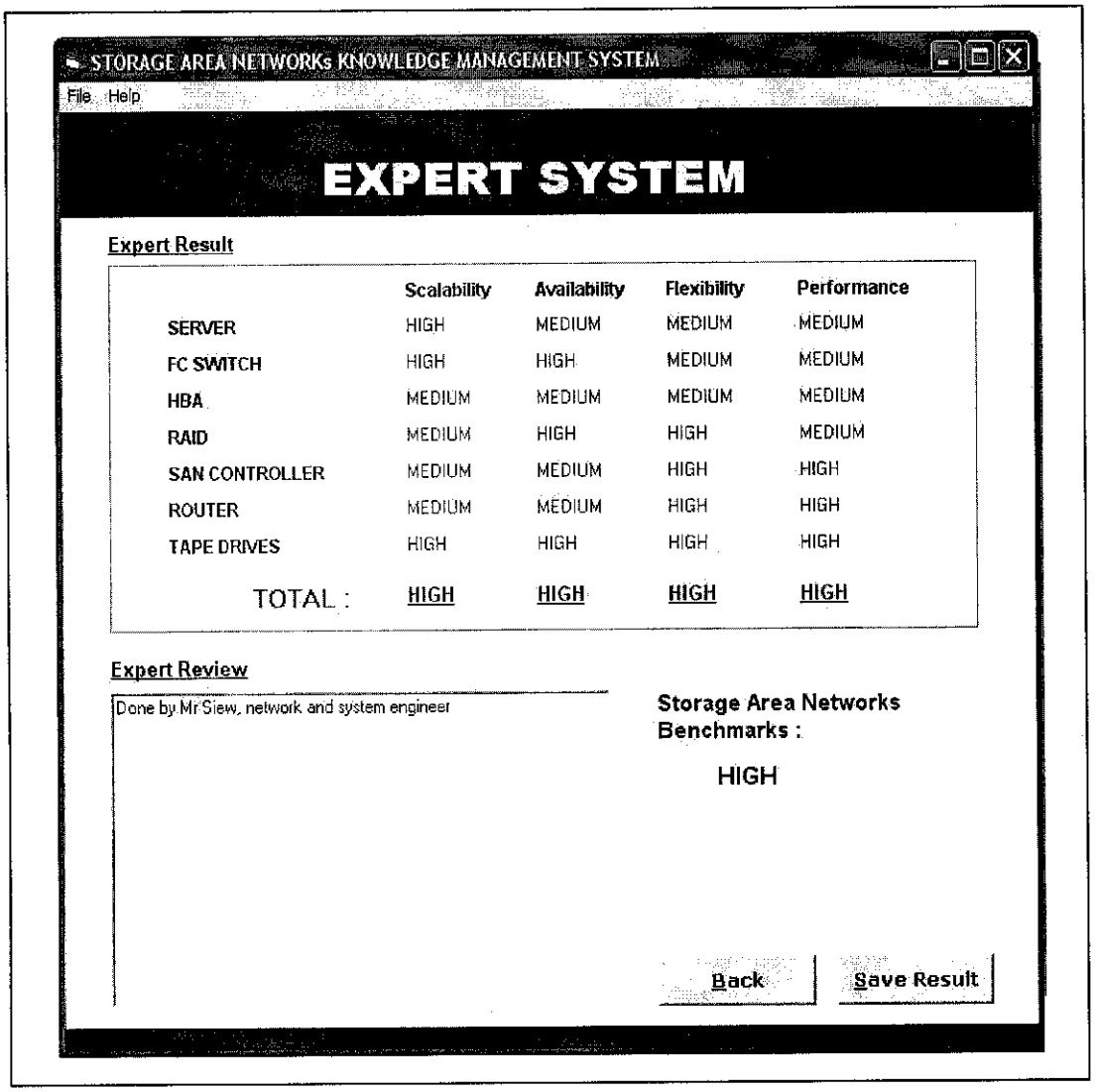


Figure 16.0 : Expert System Benchmarking Page

For high, medium and low selection will bring different points. For high will come out with 3 points, medium with 2 points and low with 1 point. For none selected benchmark either high, medium or low will bring 0 point. The overall points will be calculated based on the scalability, availability, flexibility and performance. For the high benchmark, it must 80% of the overall points in the criteria. For the medium must be from 79% till 50% of the overall points. Overall points for each criteria are 21 points and the overall point for the SANs is 84 points.

4.6.1 Artificial Intelligence Algorithm in Decision Support Function

In designing decision support system in the Storage Area Networks Knowledge Management System, Artificial Intelligence architecture is used as the basic mechanism in searching the best SANs solution for the business. Decision tree is the AI concept used in the SANs KMS. The decision tree is based on the cost estimation which the user key-in the system earlier. In SANs, seven hardware (h) have been identified to accomplish the SANs solution. Servers, fiber channel switch, host bust adapter (HBA), redundant arrays of inexpensive disks, SAN controller, router and backup drive are the hardware that used in SANs solution. From the hardware, it divided into two categories either high cost or low cost and it makes that for each hardware, there are two possibilities (p) either low cost or high cost.

Simple formula $\mu = p^h$

$p = 2 \quad h = 7$

$\mu = p^h$

$\mu = 2^7$

$\mu = 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2$

$\mu = 128$

From the equation, found that there are 128 difference solutions for the SANs. If the hardware are divided into three categories, low cost, medium cost and high cost, the possibilities would become three and the result would be difference from above.

$p = 3 \quad h = 7$

$\mu = p^h$

$\mu = 3^7$

$\mu = 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3$

$\mu = 2187$

The equation above show that if the probabilities increase, the result become bigger where there are 2187 SANs solution based on the cost estimation. This only for simple equation using $\mu = p^h$. If the probabilities are dynamic, for example the probabilities are not the same for each hardware, then the result will be totally different from the formula.

Dynamic formula $\mu = p_1^{h1} (p_2^{h2}) (p_3^{h3}) (p_i^{hi}) (p_{i+1}^{hi+1})$

Consider the following example:-

Servers – 2 possibilities
Fiber Channel Switch- 3 possibilities
Host Bust Adapter (HBA) – 4 possibilities
Redundant Arrays of Inexpensive Disks – 2 possibilities
SAN Controller – 3 possibilities
Router - 3 possibilities
Backup Drive – 2 possibilities

$\mu = 2 \times 3 \times 4 \times 2 \times 3 \times 3 \times 2$
$\mu = 864$
From the example, the dynamic possibilities give the 864 combination of SANs hardware. Figure 1.0 shows the depth-first algorithm searching graph

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 CONCLUSION

The major part in the SANs is the architecture which suitable to the business environment based on the several criteria and standard. The system that determine SANs architecture helps the business entity in designing and giving expert review in designing Storage Area Networks (SANs) that are effectively and efficiently. The implementation and development will be done using various functional artificial intelligent technology standard with supported by the Visual Basic 6 as the interface. The student will consider networking standard and expert review in designing the system procedures.

SANs can provide many benefits. Centralizing data storage operations and their management is certainly one of the chief reasons that SANs are being specified and deployed today. Administrating all of the storage resources in high-growth and mission-critical environments can be daunting and very expensive. SANs can dramatically reduce the management costs and complexity of these environments while providing significant technical advantages.

Many companies will evolve towards SAN technology without really realizing it, as new disk arrays, offline storage devices, servers and software are all made fiber and SAN connectable. Others will adopt Storage Area Networks aggressively to gain the potential benefits of far better utilization of resources, new levels of availability, performance and the ability to better share corporate data across their company. Earlier adopters of Storage Area Networks must select proven suppliers with a pedigree for producing reliable hardware or software. They must advance aggressively to gain

benefits of a SAN, but with due care and planning. Management tools and suitable SAN aware utilities should be deployed early to help control this new unknown.

Storage Area Networks will take several years to mature during which we will see these early adopters work through an evolution from SCSI on steroids to an environment with outstanding availability and performance for global applications and globally shared data. The key to success is the establishment of a storage centric environment supported by a high performance low latency fiber fabric that provides users with highly available access to clusters of application servers with many-to-many connectivity to share online and offline storage.

5.2 RECOMMENDATION

Fibre Channel Industry Association, the Storage Networking Industry Association and other organizations have pushed the technology and pooled resources to help develop a set of standards for management. However, these groups have yet to develop standards to address the lack of interoperability between switched fabric implementations. Compaq Computer Corp., EMC Corp., Hewlett-Packard Co., Sun Microsystems and other vendors have been quick to embrace the growing SAN market, offering complete solutions. Additionally, many vendors now offer a full line of products that can use to brew own SANs.

For future enhancement, the SANs KMS would be on web based system where the user can login from the internet. The web based system will enable the user to collaborate between the expert users all over the world. The users also would become more to public domain where everybody can access to the SANs KMS with several restrictions for the security purpose. Besides that, the system will require a huge and complex database system which Oracle is the best solution for the database architecture. Once the database is huge enough, the system can use dynamic searching instead of static searching. The system also will use may hardware and software to generate more solutions from the current system which use two probabilities for each hardware consideration. The higher the probabilities, the better in searching the best solution for the SANs.

REFERENCES

LITERATURE:

1. Network World, May 06, 2002 "Business Continuance"

2. SGI, "Storage Area Network"
   URL: http://www.sgi.com/products/storage

3. HP, "StorageWorks SAN Storage Area Network"
   URL: http://www.hp.com

4. Law firm Clark, Thomas & Winters of Austin TX, "Storage area networks"
   URL: http://www.tivoli.com/

5. Fiber Channel-Based Storage Area Network, July 28, 2000

6. Patuxent Technology Partners of Clarksville, 2000 "SANs for Corporation and Government Agencies"
   URL: http://www.tivoli.com/

7. Fibre Channel Industry Association, 2003
   URL: http://fibrechannel.org/

8. Mylex Corporation, 2003
   URL: http:/ www.mylex.com

9. 3Com Corporation, 2003
   URL: http://www.3com.com/moreconnectedbusiness

10. Gadzoox Networks for Storage Area Networking
URL: http://www.gadzoox.com

11. Net Age, 1999 "Looking at Real-World Convergence"

12. Building SANs with its Shark Enterprise Storage Servers, IBM
URL: http://www.ibm.com/storage

13. Tape libraries figure prominently in SANs, Storagetek
URL: http://www.storagetek.com/

14. Storage Network Industry Association,2000
URL: http://www.snia.org/

15. SAN Insight, 2000 "A comprehensive SAN management utility"
URL: http://www.vixel.com/

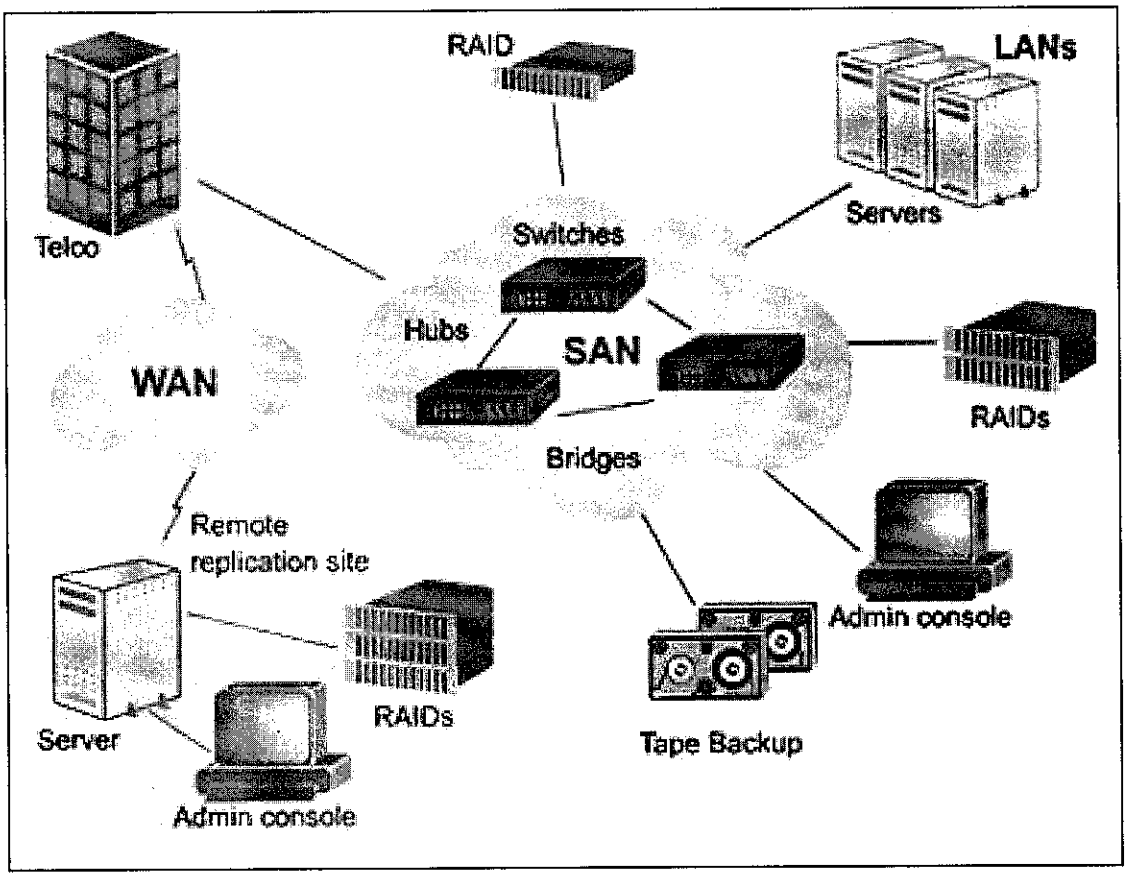


Figure 17.0 : Architecture of Storage Area Network

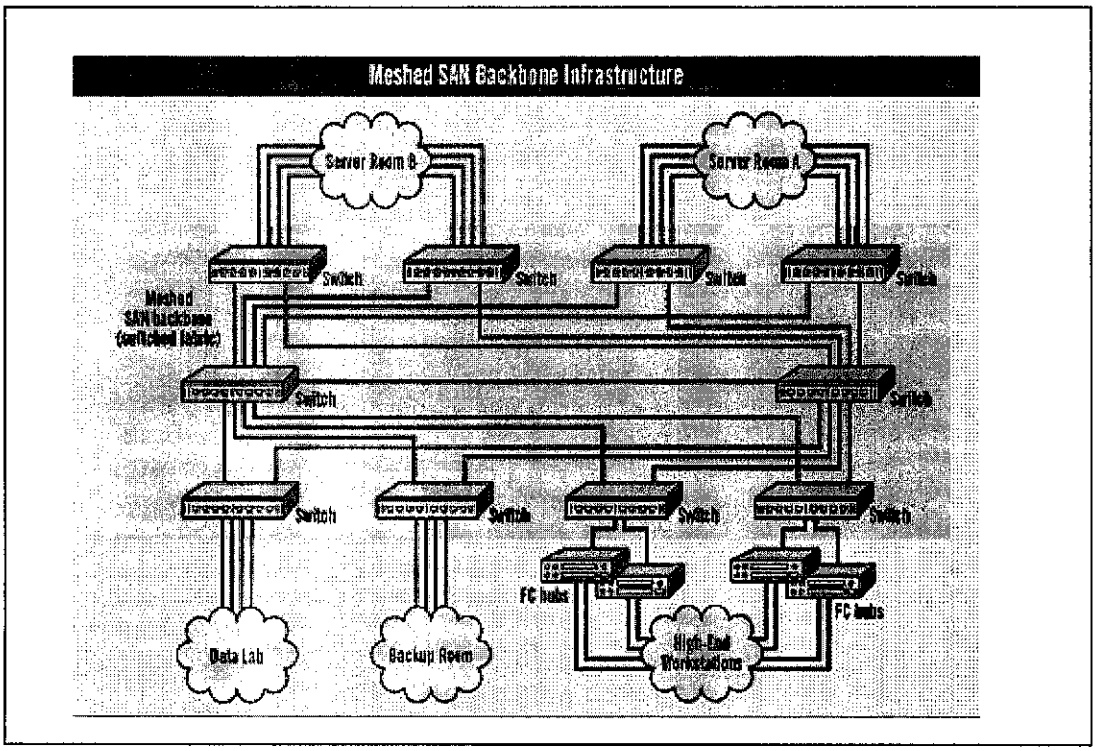


Figure 18.0 : SANs Backbone Infrastructure

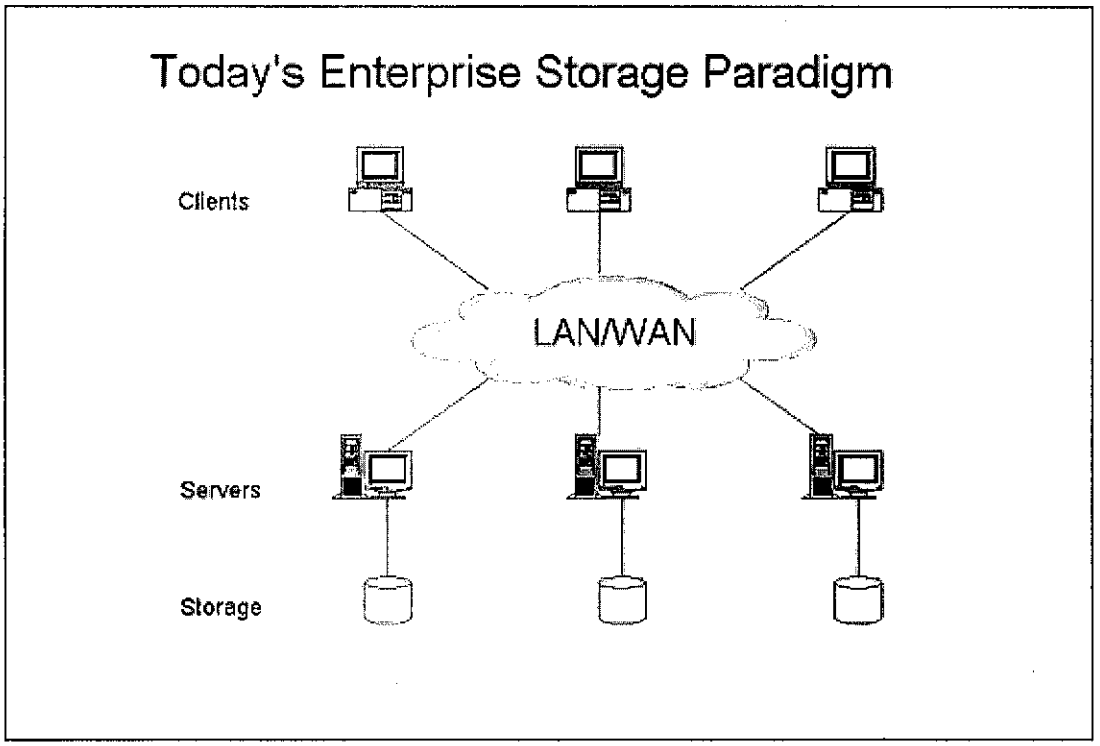


Figure 19.0 : Traditional Storage Architecture

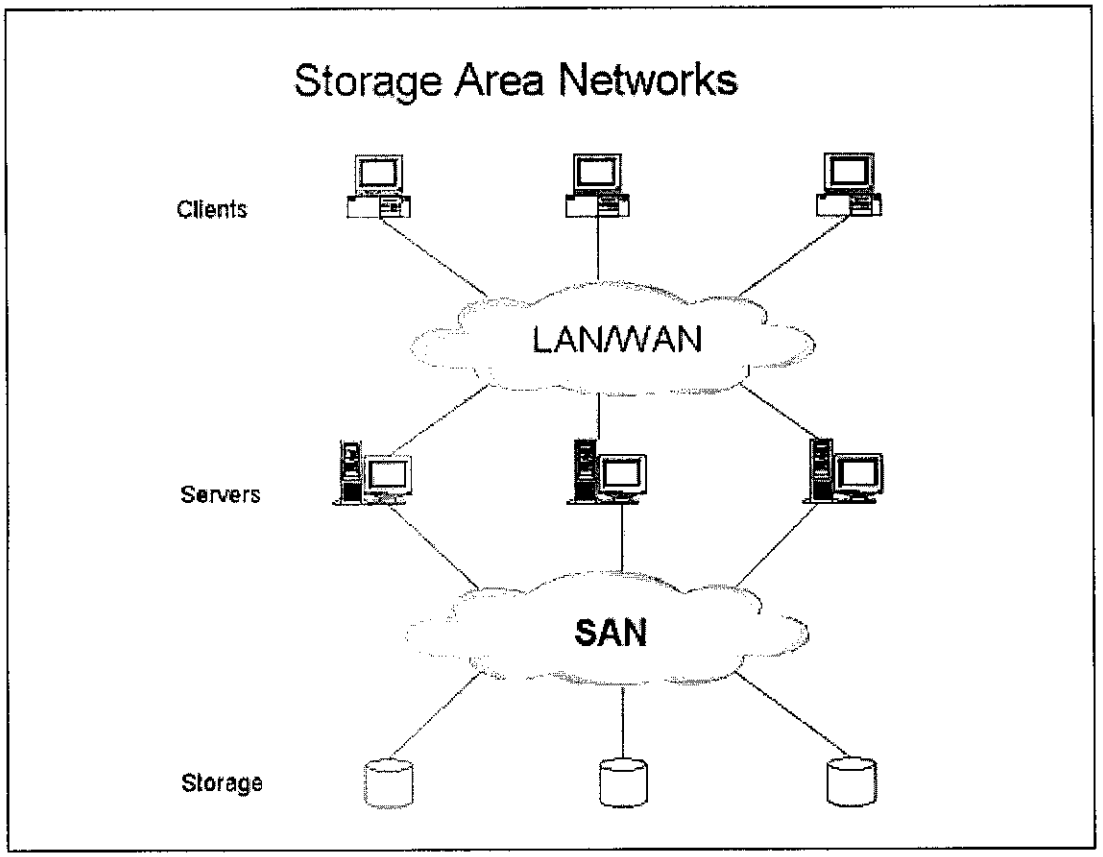


Figure 20.0 : SANs Basic Architecture

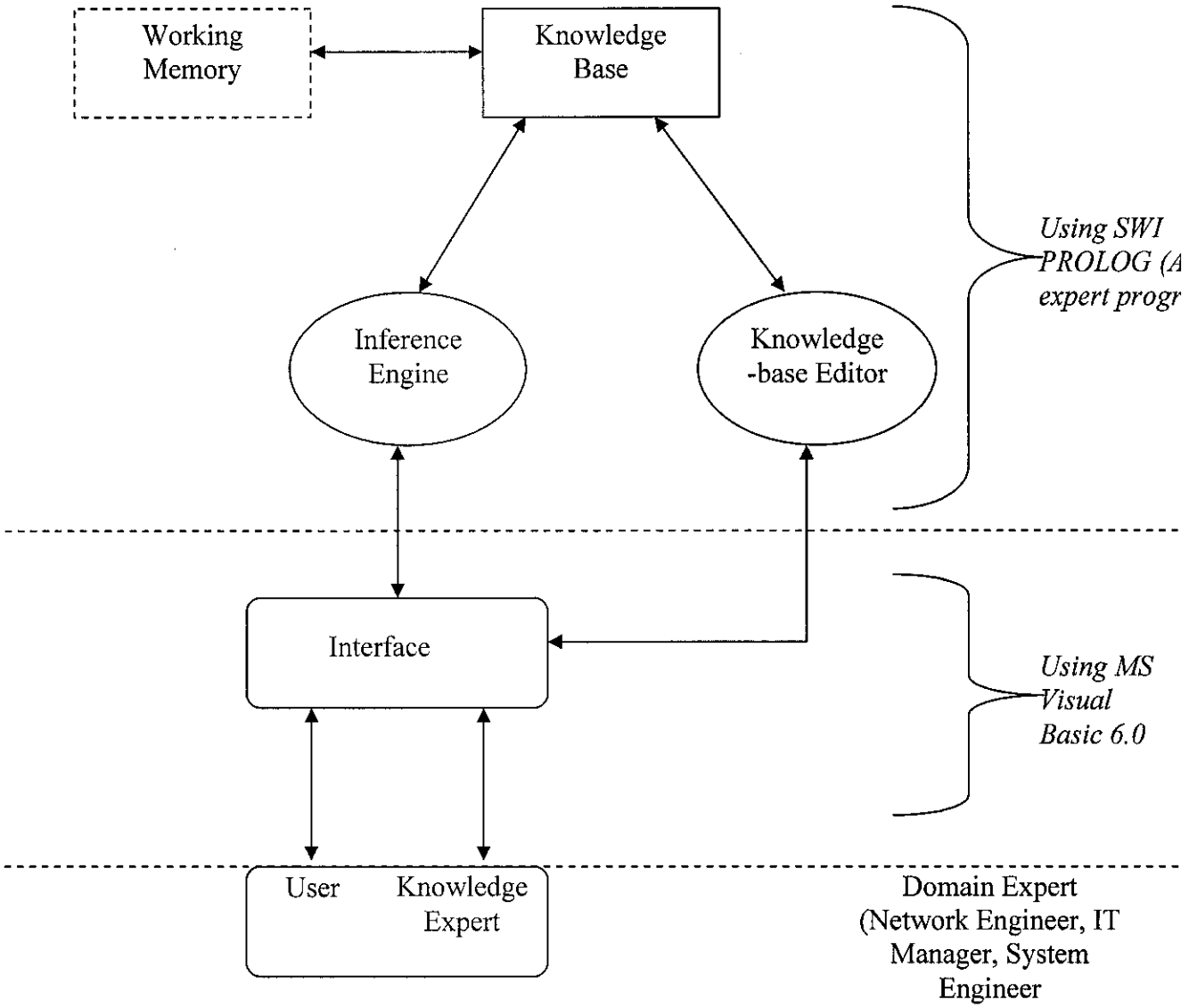


Figure 21.0 : Storage Area Networks Knowledge Management System Architecture adopted from Generic Expert System Design

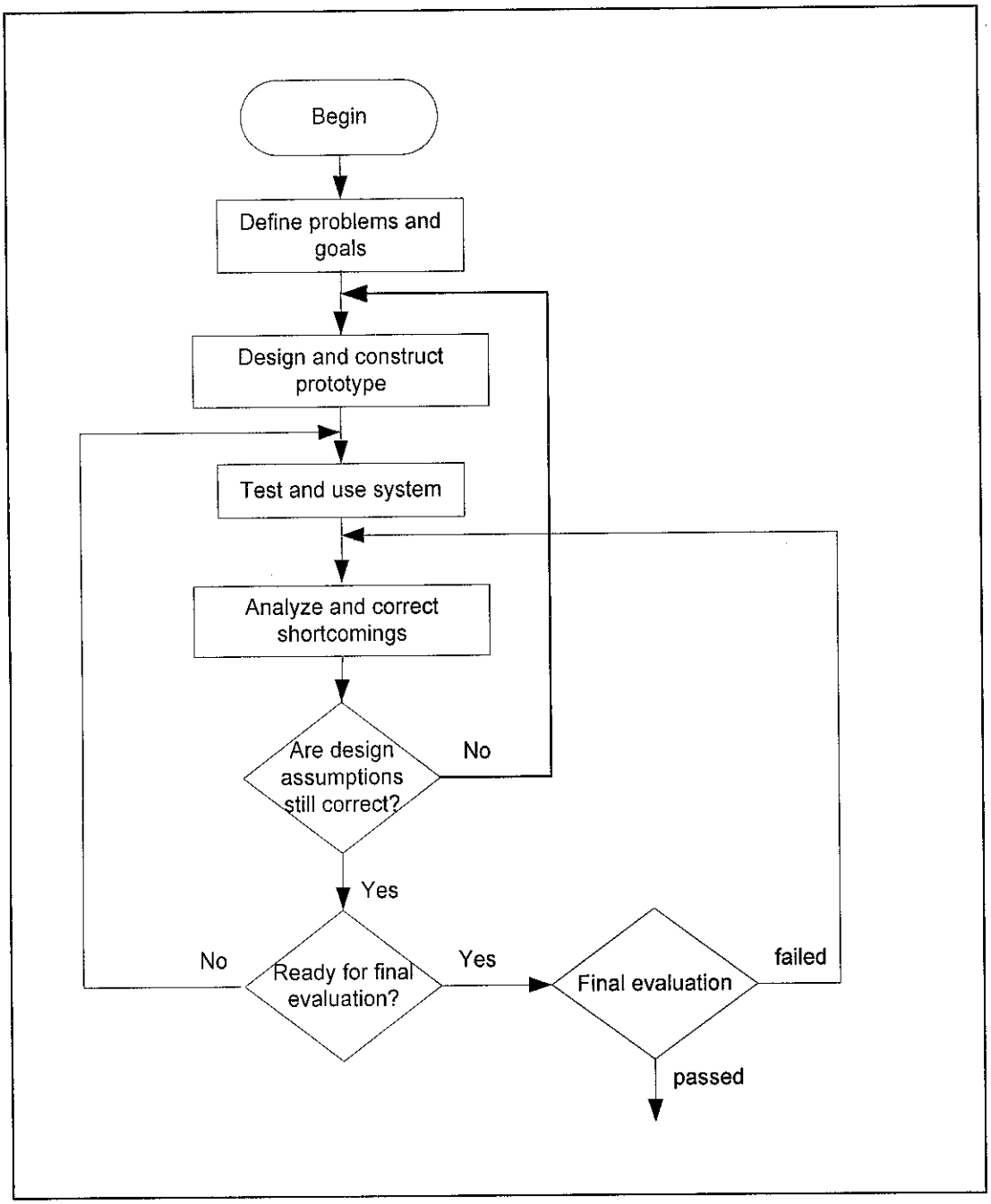


Figure 22.0 : Exploratory Development Cycle

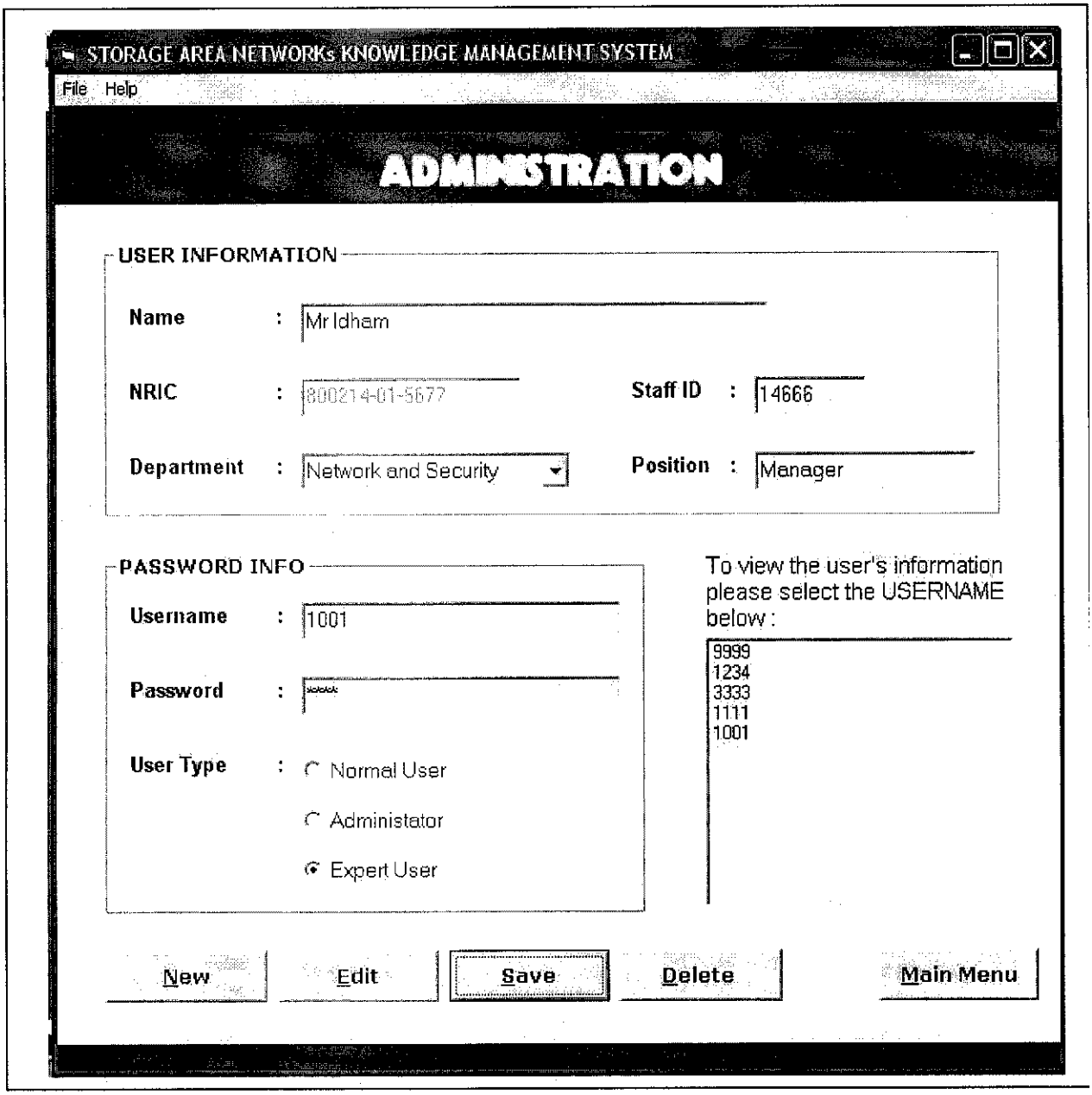


Figure 23.0 : Administration Page

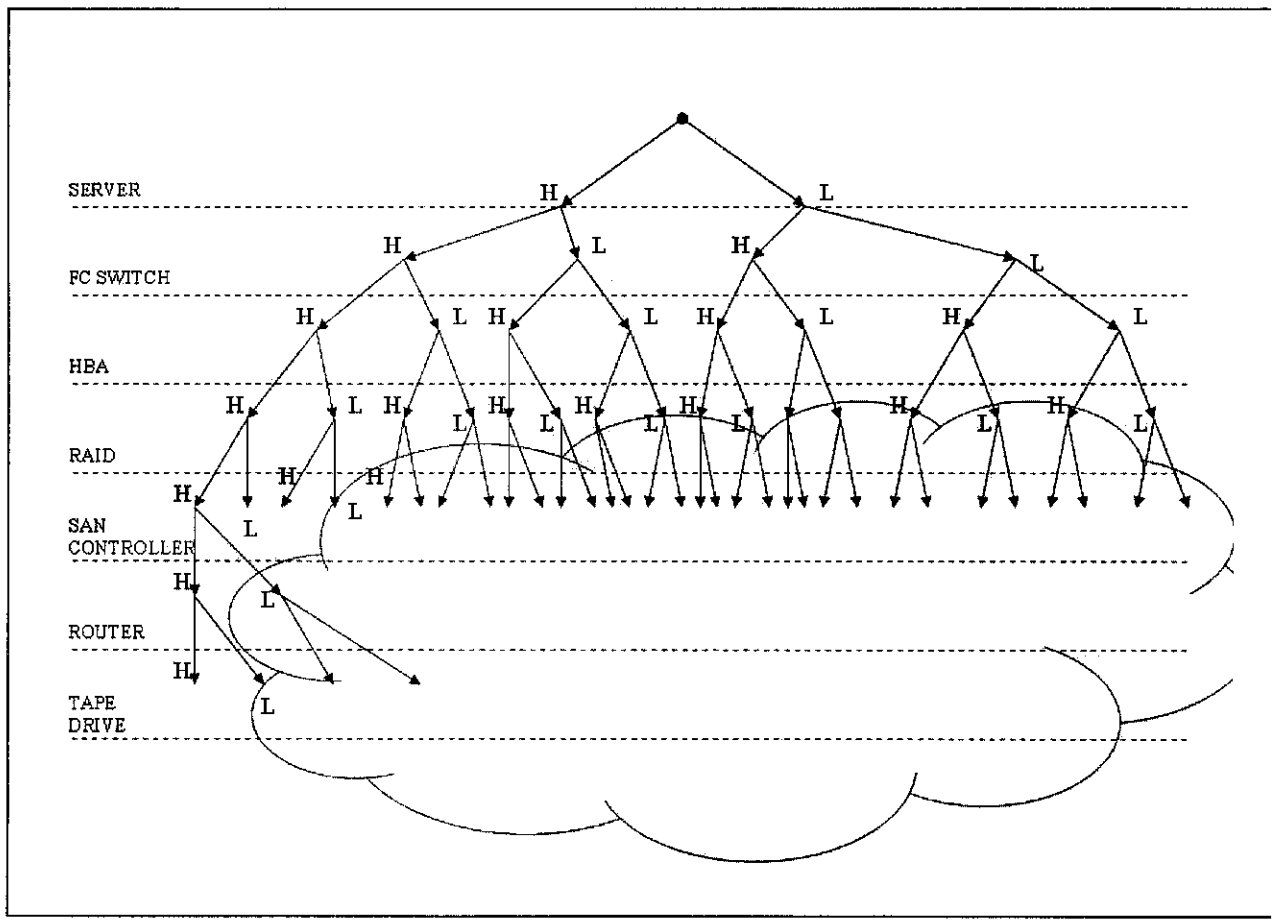


Figure 24.0 : Depth-First Searching Tree

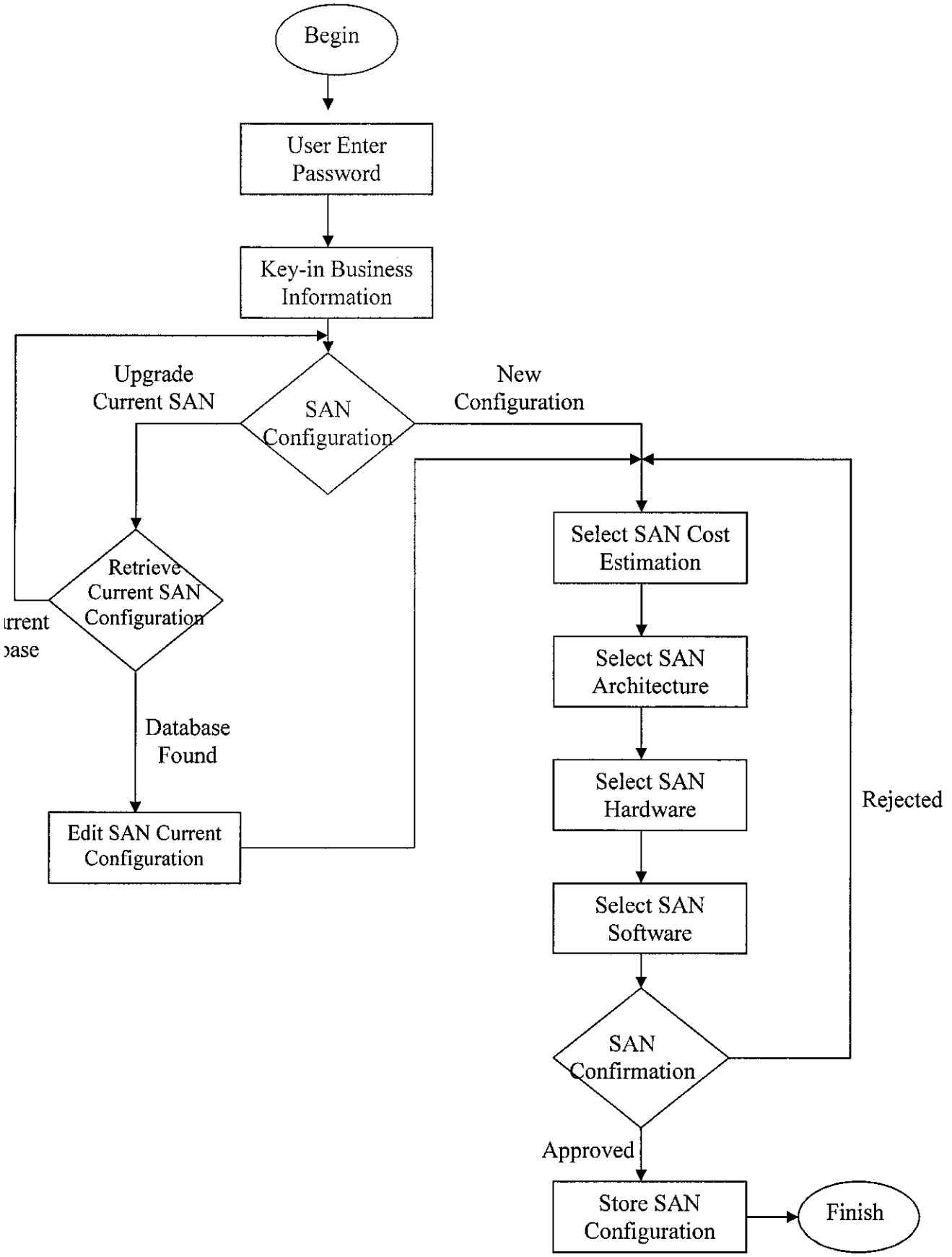
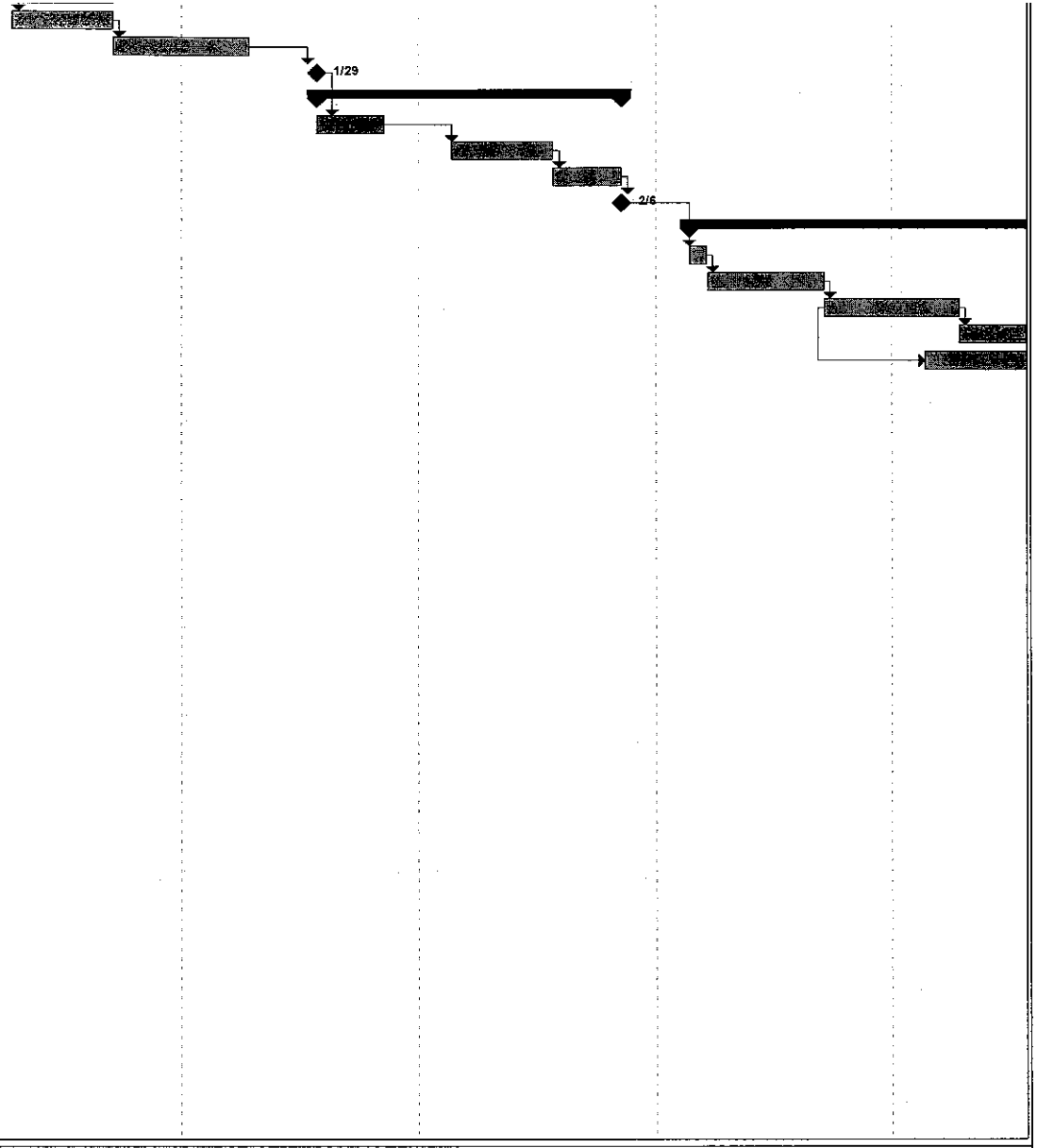


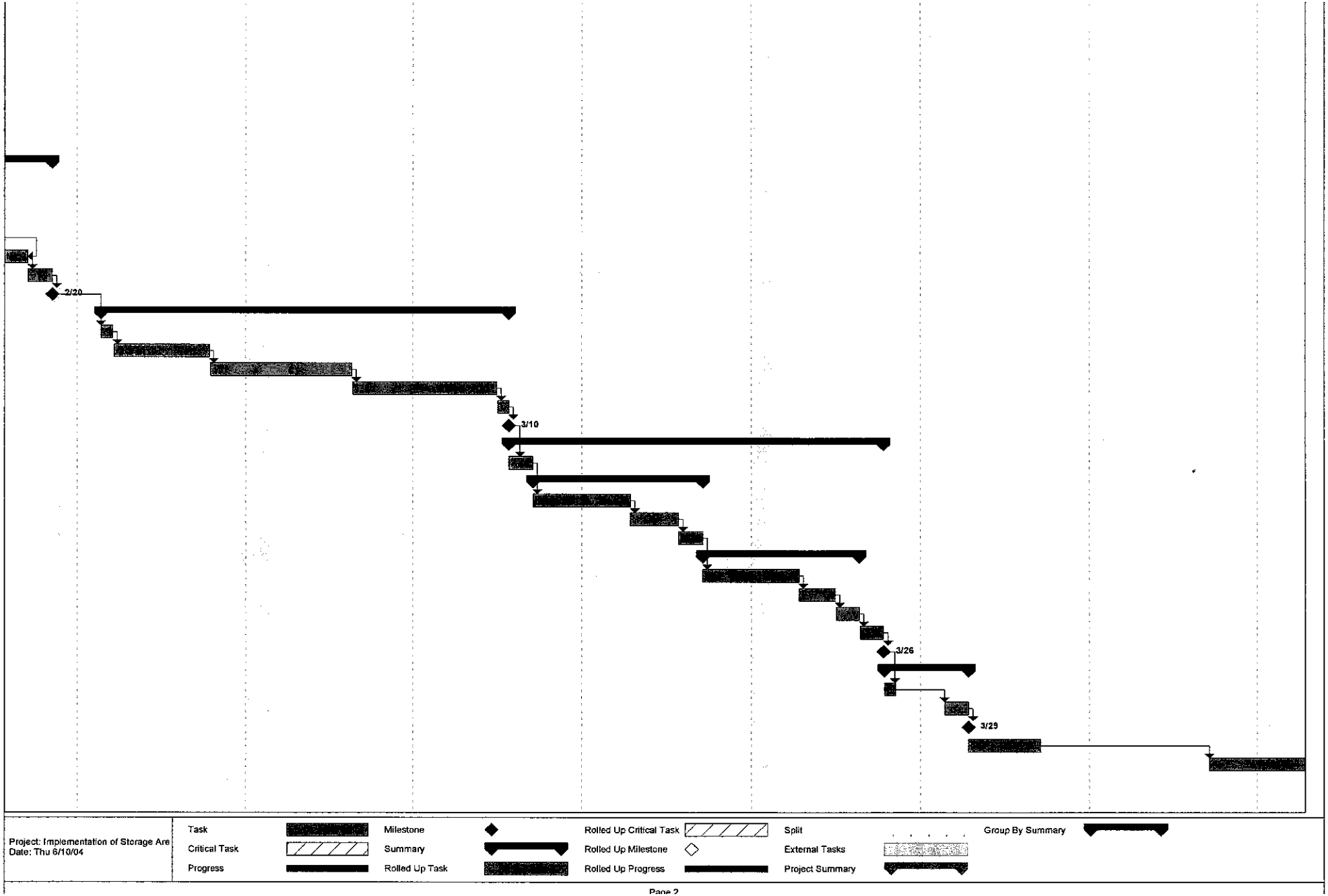
Figure 25.0 : Storage Area Networks Knowledge Management System Flow Model

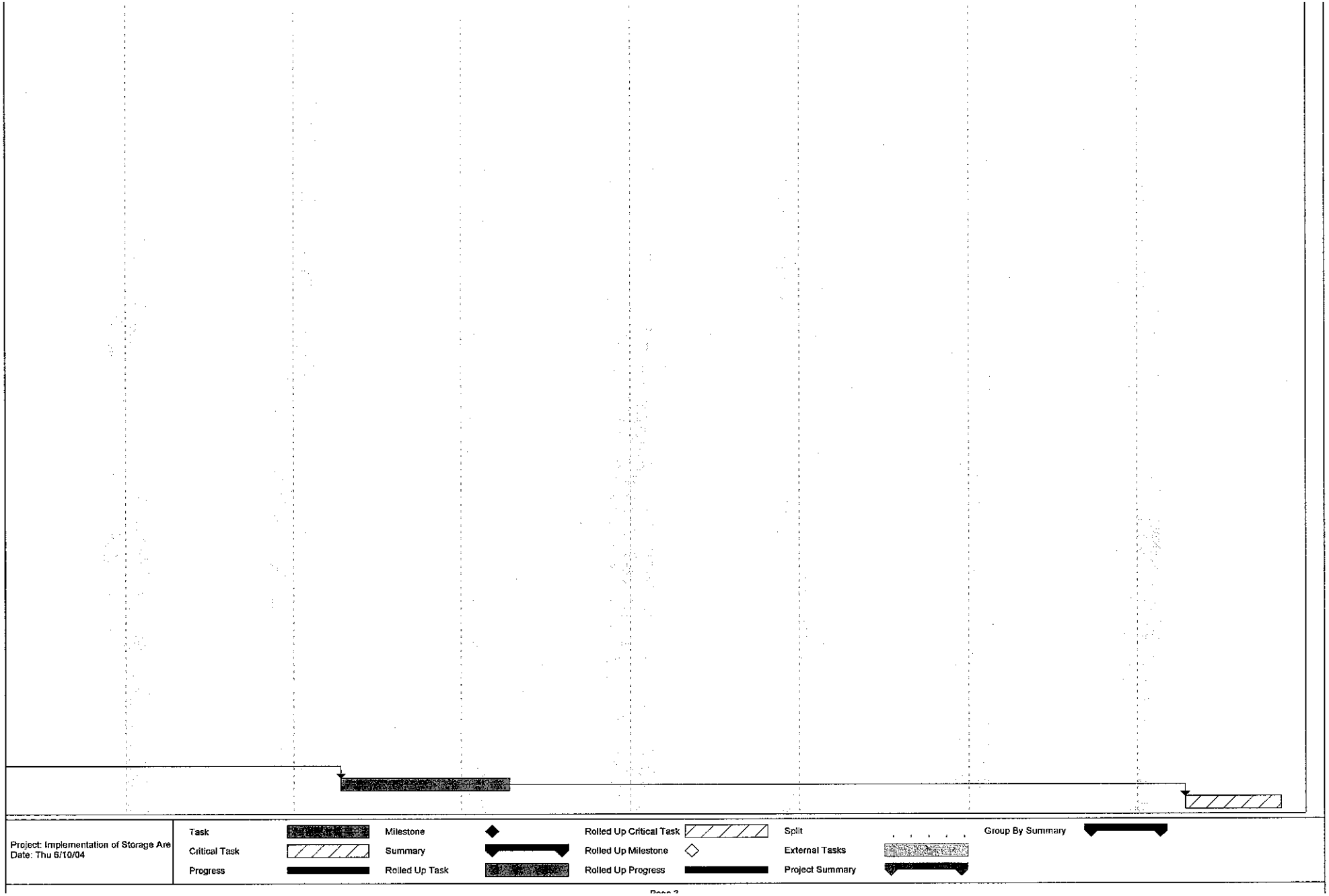
		Task Name	Duration	Start	Finish
4		Preliminary Study	3 days	Tue 1/20/04	Thu 1/22/04
5		Feasibility Study	2 days	Fri 1/23/04	Mon 1/26/04
6		Preliminary Report Submission	0 days	Thu 1/29/04	Thu 1/29/04
7		**Project Analysis**	**7 days**	**Thu 1/29/04**	**Fri 2/6/04**
8		Problem Analysis	2 days	Thu 1/29/04	Fri 1/30/04
9		Requirement Analysis and Specification	3 days	Mon 2/2/04	Wed 2/4/04
10		Data, Process and Object Modelling	2 days	Thu 2/5/04	Fri 2/6/04
11		System Requirement Documentation	0 days	Fri 2/6/04	Fri 2/6/04
12		**Project design**	**10 days**	**Mon 2/9/04**	**Fri 2/20/04**
13		Review System Requirement	0.5 days	Mon 2/9/04	Mon 2/9/04
14		Architecture Design	3.5 days	Mon 2/9/04	Thu 2/12/04
15		Interface Design	2 days	Fri 2/13/04	Mon 2/16/04
16		Database Design	2 days	Tue 2/17/04	Wed 2/18/04
17		Testing and Debugging	4 days	Mon 2/16/04	Thu 2/19/04
18		Design and Requirement Review	1 day	Fri 2/20/04	Fri 2/20/04
19		Detailed Design Complete	0 days	Fri 2/20/04	Fri 2/20/04
20		**Project Development**	**13 days**	**Mon 2/23/04**	**Wed 3/10/04**
21		Review System Design	0.5 days	Mon 2/23/04	Mon 2/23/04
22		Interface Programming	4 days	Mon 2/23/04	Fri 2/27/04
23		Database and Administrative Function	4 days	Fri 2/27/04	Thu 3/4/04
24		Developer Testing and Debugging	4 days	Thu 3/4/04	Wed 3/10/04
25		Development and design Review	0.5 days	Wed 3/10/04	Wed 3/10/04
26		System Development Complete	0 days	Wed 3/10/04	Wed 3/10/04
27		**Project testing**	**11.5 days**	**Thu 3/11/04**	**Fri 3/26/04**
28		Develop System Test Plan using System Requirement ar	1 day	Thu 3/11/04	Thu 3/11/04
29		**Unit Testing**	**5 days**	**Fri 3/12/04**	**Thu 3/18/04**
30		Interface Testing	2 days	Fri 3/12/04	Mon 3/15/04
31		Database and Administrative Testing	2 days	Tue 3/16/04	Wed 3/17/04
32		Modify and Re-test Design and Element	1 day	Thu 3/18/04	Thu 3/18/04
33		**Integration Testing**	**4.5 days**	**Fri 3/19/04**	**Thu 3/25/04**
34	✓	System Integration Testing	2 days	Fri 3/19/04	Mon 3/22/04
35		Modify System Integration	1.5 days	Tue 3/23/04	Wed 3/24/04
36		Re-test Integration Testing	1 day	Wed 3/24/04	Thu 3/25/04
37		Amendment	1 day	Thu 3/25/04	Fri 3/26/04
38		Complete System	0 days	Fri 3/26/04	Fri 3/26/04
39		**System Implementation**	**1.5 days**	**Fri 3/26/04**	**Mon 3/29/04**
40		Develop User Manuals	0.5 days	Fri 3/26/04	Fri 3/26/04
41		System Roleout	1 day	Mon 3/29/04	Mon 3/29/04
42		System Decommissionning	0 days	Mon 3/29/04	Mon 3/29/04
43		**Submission of Supervisor's Final Draft**	3 days	Tue 3/30/04	Thu 4/1/04
44		**Submission of Project's Final Draft**	2 days	Fri 4/9/04	Mon 4/12/04
45		**Oral Presentation**	6 days	Tue 4/27/04	Mon 5/3/04
46		**Submission of Project Dissertation**	4 days?	Tue 6/1/04	Fri 6/4/04



Project: Implementation of Storage Are
Date: Thu 6/10/04

Task
Critical Task
Progress
Milestone
Summary
Rolled Up Task
Rolled Up Critical Task
Rolled Up Milestone
Rolled Up Progress
Split
External Tasks
Project Summary
Group By Summary


2/20
3/10
3/26
3/29
Project: Implementation of Storage Are
Date: Thu 6/10/04
Task
Critical Task
Progress
Milestone
Summary
Rolled Up Task
Rolled Up Critical Task
Rolled Up Milestone
Rolled Up Progress
Split
External Tasks
Project Summary
Group By Summary


Project: Implementation of Storage Are
Date: Thu 6/10/04
Task
Critical Task
Progress
Milestone
Summary
Rolled Up Task
Rolled Up Critical Task
Rolled Up Milestone
Rolled Up Progress
Split
External Tasks
Project Summary
Group By Summary