Effective Data Backup System Using Storage Area Network Solution

Ume, L.E.
Department of Computer Science, Ebonyi State University, Abakaliki, Ebonyi State, Nigeria
Correspondence e-mail: leo.manuels@gmail.com

Abstract

One of the most crucial benefits of the computer system is its ability to manage data send to it for processing. An aspect of data management that all other aspects hinge upon is data storage and retrieval. Data loss has rendered many firms bankrupt. The primary cause of data loss is lack or non- existent of data backup. Storage Area Network Solution (SANS) is internet-based software which will collect clients data and host them in several locations to forestall data loss in case of disaster in one location. The researcher used adobe Dreamweaver (CSC3) embedded with PHP incorporated with MySQL database technology to develop the application. The objectives of the research were realized.

Keywords: Backup, Storage area network, Data, Effective and Data loss

---

1.0 Introduction

Storage Area Network (SAN) is a high-speed sub-network of shared storage devices. A storage device is a machine that contains nothing but a disk or disks for storing data. A SAN's architecture works in a way that makes all storage devices available to all servers on a LAN or WAN. As more storage devices are added to a SAN, they too will be accessible from any server in the larger network. In this case, the server merely acts a pathway between the end user and the stored data. A storage area network can use existing communication technology such as IBM's optical fibre ESCON or use fibre channel technology. Some SAN system integrators liken it to the common storage bus (flow of data) in a personal computer that is shared by different kinds of storage devices such as a hard disk or a CD-ROM player [6].

Storage area network support disk mirroring, backup and restore, archival and retrieval of archived data, data migration from one storage of data among different servers in a network. SANs can incorporate sub-networks with network-attached storage (NAS) systems. SAN evolved from the concept of taking storage devices and therefore storage traffic, off the LAN and creating a separate back-end network designed specifically for data. This is a contrast to the use of a traditional LAN for providing a connection for server-storage, a strategy that limits overall network bandwidth. SANs address the bandwidth bottlenecks associated with LAN based server storage and the scalability limitations found with SCSI bus based implementations. SANs provide modular scalability, high-availability, increased fault tolerance and centralized storage management. These advantages have led to an increase in the popularity of SANs as they are quite simply better suited to address the data storage needs of today's data intensive network environments. Because stored data does not reside directly on any of a network's servers. Server power is utilized for business applications and network capacity is released to the end user.

1.1 Objective Of The Research

The volume of data generated by the use of Information and Communication Technology (ICT) tools to carry our routine duties has made adequate back-up and good storage technology an imperative. They have been some existing innovation in this case which SAN is one of such that this study tends to critically analyze.

2.0 Literature Review

According to [8] describes a storage area network as similar to Local Area Network (LAN), but designed to handle large data transfers. A SAN typically supports data storage, retrieval and replication on business networks using high-end servers, multiple disk arrays fibre channel interconnection technology. Storage Area Network technology is similar but distinct from network attached storage (NAS) technology. While SANs traditionally employ low-level network protocols for transferring disk blocks, a Network Attached Storage device typically works over TCP/IP and can be integrated fairly easily into home computer networks.

Historically,[1] stated that data centers first created "Islands" of Small Computer System Interface (SCSI) disk arrays as direct-attached storage (DAS), each dedicated to an application, and visible as a number of "virtual hard drives" (i.e. LUNs). Essentially, a SAN consolidates such storage islands together using a high-speed network.

Despite such issues, SANs help to increase storage capacity utilization, since multiple servers consolidate their private storage space onto the disk arrays.

Common uses of a SAN include provision of transactional accessed data that require high-speed block-level access to the hard drives such as email servers, databases and high usage file servers.

Storage Area Networks also tend to enable more effective disaster recovery processes. A SAN could span a distant location containing a secondary storage array. This enables storage replication either implemented by disk array controllers by server software or by specialized SAN devices. Since IP WANs are often the least costly method of long-distance transport, the fibre channel over IP (FCIP) and iSCSI protocols have been developed to allow SAN extension over IP networks. The traditional physical SCSI layer could only support a few meters of distance not nearly enough to ensure business continuance in a disaster [3].

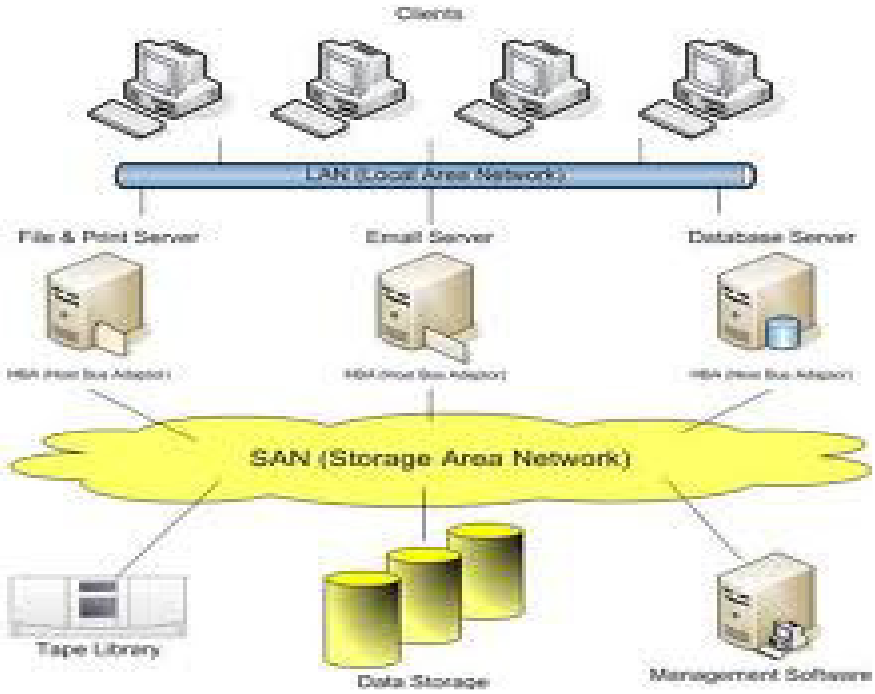

Fig 1: Architecture of SAN

Sources: [2]

Key Considerations in developing a Storage Area Network Design

Storage Area Networks (SANs) can let several servers share storage resource and are often used in situations that require high performance or shared storage with block-level access, like virtualized servers and clustered databases [4].

Although SANs started out as a high-end technology used only in large enterprises, cheaper SANs are now affordable even for small and medium-sized business (SMBs).

Uptime and availability

Because several servers will rely on a SAN for all of their data, it's important to make the system very reliable and eliminate any single points of failure. In a typical storage area network design, each storage device connects to a switch that then connects to the servers that need to access the data. If either paths fails, software can fail over to the other. Some programs will handle that failover automatically, but cheaper software may require you to enable the failover manually.

Capability and scalability

A good storage area network design should not only accommodate your client's current storage needs, but it should also be scalable so that your client can upgrade the SAN as needed throughout the expected lifespan of the system. You should consider how scalable the SAN is in terms of storage capacity, number of devices it supports and speed.

One feature of storage area network design that you should consider is thin provisioning of storage. Thin provisioning tricks servers into thinking a given volume within a SAN, known as a Logical Unit Number (LUN) has more space than it physically does. But this storage area network design requires more maintenance down the road, it's best for stable environments where a client can fairly accurately predict how each LUN's storage needs will grow.

Security

With several servers able to share the same physical hardware, it should be no surprise that security plays an important role in a storage area network design. Most of this security work is done at the SAN's switch level; zoning allows you to give only specific servers' access to certain LUNs, much as a firewall allows communication on specific ports for a given IP address. If any outward facing application needs to access the SAN, like a website, you should configure the switch so that only that server's IP address can access it.

Replication and disaster recovery

With so much data stored on a SAN, your client will likely want you to build disaster recovery into the system. SANs can be set up to automatically mirror data to another site, which could be a failsafe SAN a few meters away or a disaster recovery (DR) site hundreds of thousands of miles away. If the client wants to build mirroring into the storage area network design, one of the first considerations is whether to replicate synchronously or asynchronously.

3.0 Research Methodoly and System Analysis

A methodology is a formalized approach to implementing the System Development Life Cycle (SDLC) that is, is a list of steps and deliverables. There are many different systems development methodologies and each one is unique because of its emphasis on processes versus data and the order and focus it places on each SDLC phase. There are different types of methodologies and the emphasis they place on processes versus data and the order and focus they place on each system development life cycle. Some of the types of methodologies are discuss below:

Structured System Analysis and Design Method (SSADM)

SSADM provides a systematic approach to analysis and design of information technology (IT) applications. SSADM covers those aspects of the life-cycle of a system from the feasibility study stage to the production of a physical design; it is generally used in conjunction with other methods, which is concerned with the broader aspects of project management.
In detail, SSADM sets out a cascade or waterfall view of systems development, in which there are a series of steps, each of which leads to the next step. (This might be contrasted with the rapid application development - RAD - method, which pre-supposes a need to conduct steps in parallel.

Object-oriented analysis and design (OOAD)

This is a popular technical approach to analyzing, designing an application, system, or business by applying the object-oriented paradigm and visual modeling throughout the development life cycles to foster better stakeholder communication and product quality.

The purpose of any analysis activity in the software life-cycle is to create a model of the system's functional requirements that is independent of implementation constraints.

The main difference between object-oriented analysis and other forms of analysis is that by the object-oriented approach we organize requirements around objects, which integrate both behaviors (processes) and states (data) modeled after real world objects that the system interacts with.

Computer-aided software engineering (CASE)

The use of a computer-assisted method to organize and control the development of software, especially on large, complex projects involving many software components and people. Using CASE allows designers, code writers, testers, planners, and managers to share a common view of where a project stands at each stage of development. CASE helps ensure a disciplined, check-pointed process. A CASE tool may portray progress (or lack of it) graphically. It may also serve as a repository for or be linked to document and program libraries containing the project's business plans, design requirements, design specifications, detailed code specifications, the code units, test cases and results, and marketing and service plans.

Some of the benefits of CASE and similar approaches are that, by making the customer part of the process (through market analysis and focus groups, for example), a product is more likely to meet real-world requirements. Because the development process emphasizes testing and redesign, the cost of servicing a product over its lifetime can be reduced considerably. An organized approach to development encourages code and design reuse, reducing costs and improving quality. Finally, quality products tend to improve a corporation's image, providing a competitive advantage in the marketplace.

Analysis of the proposed system

Before the development of the new system, an intensive analysis of the relevance of the system was carried out to ascertain the viability of the research. After analyzing the current method of backing up files in my locality (Abakaliki urban in Nigeria) I discover that there is need to develop an application that will enable schools and company or industrial sector create a storage area network that will enable them backup of their files to avoid the danger of data loss. The proposed system is designed in such a way that users will register with the system in other to acquire username and password, immediately after registration they will automatically be allotted to a memory space. This will be proceeded by payment, and then they can be login in subsequently to upload their files and sending their files to the admin for backup. The users can send a message across to the admin in request for a particular file that has been backed up and the admin will cross check and reply the users. The admin can also view all register

users, edit their record, and update their

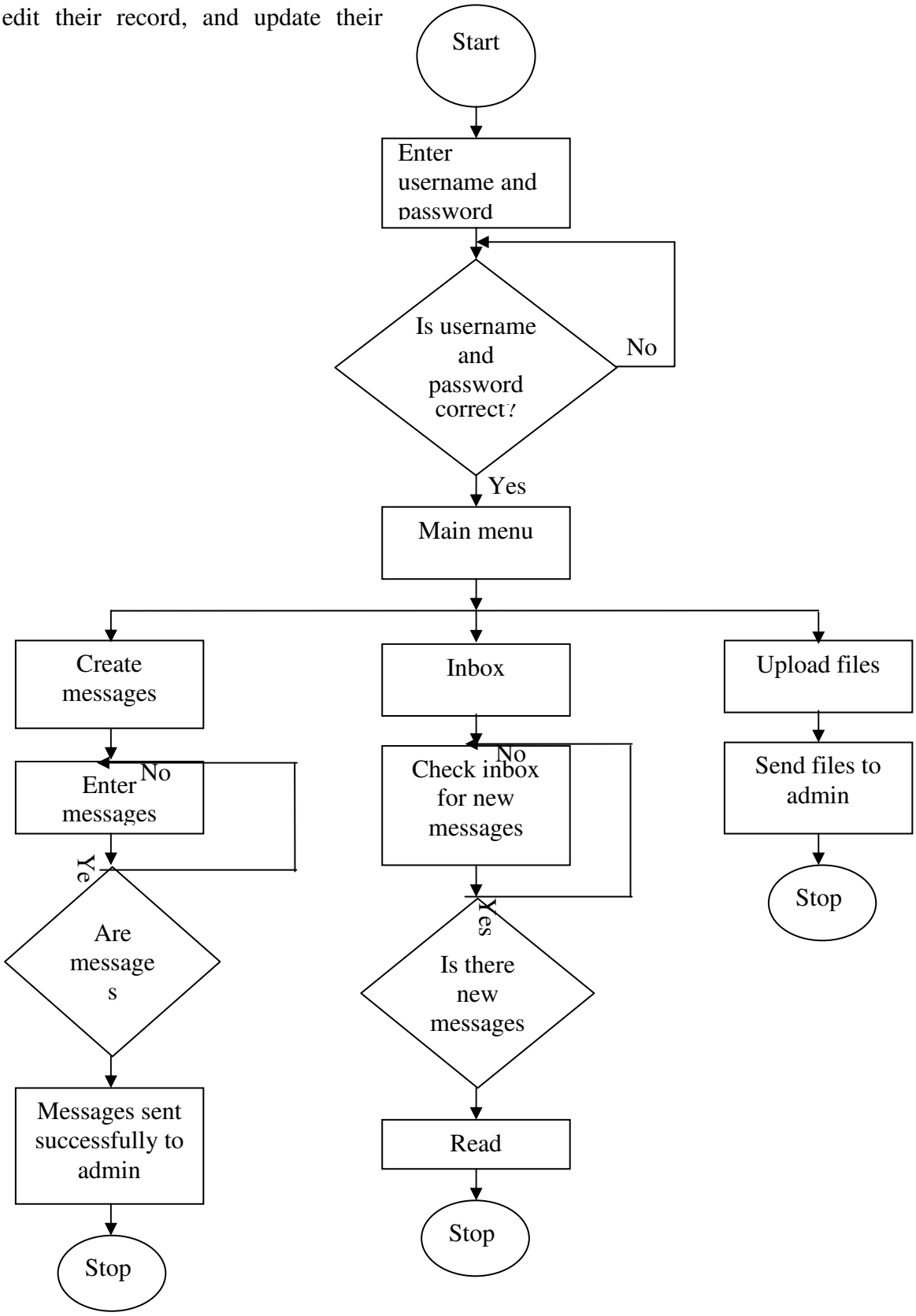

Fig 2: Flowchart of the proposed system

Strengths of the Proposed System

- **Easy access to data:** Data backed up in Storage Area Network (SAN) solution system is easier to be accessed as the infrastructures used are of high specifications.

- **Reliability:** The solution bring about some reliability. This is because SAN solution system provides a platform to back-up data in many locations. Therefore a damage in one location will not completely destroy the data as it will be near impossible for disaster to occur at all the locations where the data is back-up at the same time.

- **Security:** Data is live wire of any organization and therefore need to be properly secured. SAN solution system is designed with optimum security in sight. Using the system ensures proper security of data backed up.
- **Consistency:** This is the ability of data to retain its quality in various locations where the data is backed-up. Data backed up in the system is always the same in various locations ensuring a high integrity for the data.
- **Effective data management:** SAN solution system provides a proper management of data compared to other system.

4.0 System Specification, Design and Implementation

System Specification

Specification is literally the discussion of a specific point or issue. A project's specifications consist of the body of information that should guide the project developers, engineers and designers through the work of creating the software given by the user of the software. Hence,

This work is aimed at designing a storage area network solution system that will enable schools and cooperate organization create a backup of their files in different locations at the same time.

Input Specification and design

The input specification of this system specifies the data the user has to supply for the system to work efficiently and generate accurate results. There are three major points of data input; input is made while trying to

Login into the system, input is made when the admin register's schools or companies, and input is also made when users tries to upload files for backing up. Some this input forms are displays below:

Table 1: login form input

FIELD NAME	FIELD TYPE	FIELD LENGTH
Username	Text	15
Password	Text	9

Table 2: User Registration Input Form

FIELD NAME	FIELD TYPE	FIELD LENGTH
Company_name	Varchar	30
Company_address	Varchar	50
email	Varchar	30
Contact_number	number	13
Username	Varchar	15
Password	Varchar	9

Fig 3: Input Form design

Output Specification

The output specification displays result after the keyed data have been processed. This comprises the output to the above input, output is made after the company or school has been registered, the admin can view the record as output, output is also made when a message or file is send to the admin by the users, the admin can also view the file as output.

Database Specification

Database specification specifies the structure of the database(s) that will be used in the system. One database was created and eight tables the name of the database created is backup, MYSQL (My structured query language) was used for the creation of the database; some of the tables used are displayed below:

:

Table 4: BACKUP TABLE

S/NO	FIELD NAME	DATA TYPE	FIELD SIZE
1	Title	VARCHAR	50
2	Description	Varchar	50
3	File_upload	VARCHAR	90
4	Sender	VARCHAR	50
5	Data_uploaded	VARCHAR	20
6	Backup_status	VARCHAR	20
7	Downloaded_status	VARCHAR	10

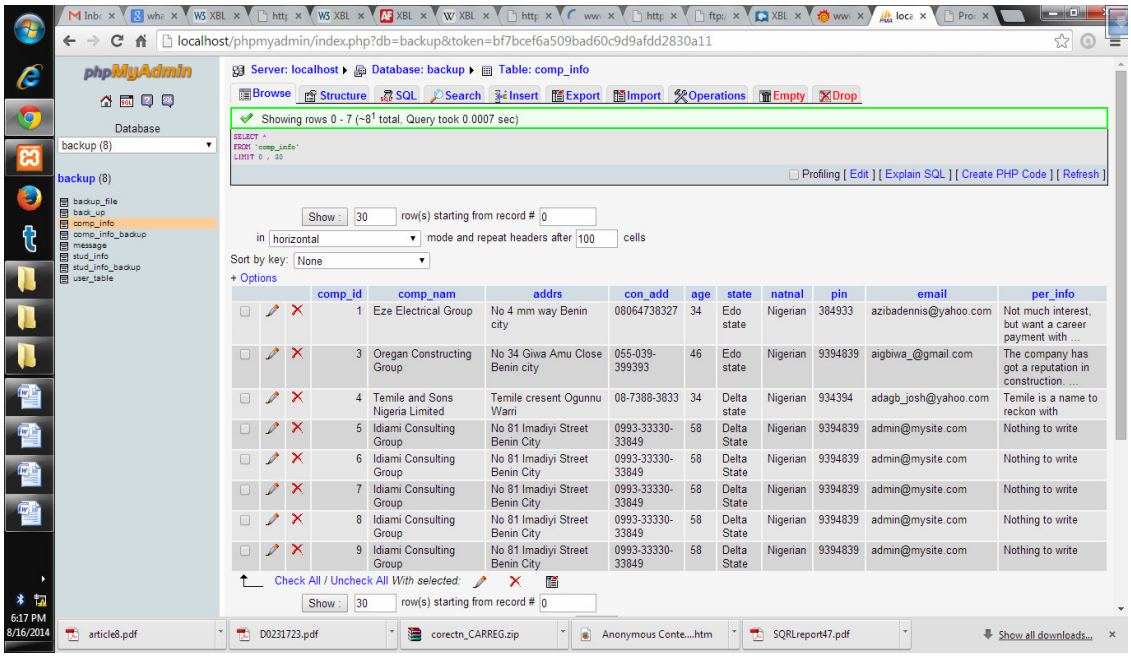

Fig. 4: Back –up Data Output

Process design

Storage Area Network solution system as a web-based application was implemented using web based development tools. There are different types of web-based development tool, but adobe macromedia dream weaver CS3 was selected for the development of the new system because other languages like scripting languages are embedded in it such as PhP (hypertext preprocessor), javascript, html (hypertext markup language) etc. and also have the following advantages:

i. it is user friendly
ii. easy to understand
iii. all other scripting language, server side scripting language are embedded which simplifies the creation of powerful and attractive web pages

A virtual server was also used in the development of the new system, xampp (cross apache, mysql, perl and php) was choosing as the virtual server, it was choosing because it has a good graphical user interface, is user friendly and it can work on any operating system

Software testing

This is the process of performing variety of tests on a system to explore functionality or identify problems. Software testing is actually required before and after the software is put into place. testing process identifies program defects or weakness. this testing was thorough to ensure that the system meets organizational and end user requirements. the coding of this project was put through series of test ranging from unit testing, integration testing, system testing and user acceptance testing.

A top down approach was used in testing this system, in which the bigger problems were first broken into smaller problems which made testing and debugging easier.

Test plan

A test plan documents the strategy that will be used to verify and ensure that a system meets its design specifications and other requirements.

The approach used in the test plan of this software is the top down method. this involves the systematic procedure of testing the main menu system first, the subsystems and finally, the different modules. hence, the top down is effective as it describes the functionality at a more detailed level.

Having designed the system using this approach, it is essential that its test plan follow the same process. the top down test

plan, which has a systematic procedure, involves the following:

i. Testing the main menu system which enables one to select an appropriate subsystem of his/her choice.

ii. Secondly, the subsystems that co-ordinate the software packages are tested and they enable the user to select a program computer program, which also involves, feeding the system with raw data to prove the reliability of the system modules, simulated data and actual data were used as test data to ensure the system is working as expected and meets required conditions

Test Result (Expected versus Actual)

This is the result obtained after testing the system with the test plan and test data.

module under that subsystem. Also, the software modules for each subsystem are tested.

Test data

Test data is data which has been specifically identified for use in tests, typically of a computer system.

During the testing, the actual and expected results were compared to ensure they produced same result or if there is a difference, it should be slight and negligible. Hence the result:

Table 5: Comparison between expected and actual results

S/N	TEST DATA	EXPECTED RESULT	ACTUAL RESULT
1	Users click on register new.	Users should be able to register with the system in other for the user to gain access to the system.	Users were able to register with the system.
2	Users login with correct username and password	Users should be granted access if username and password is correct, but if users enter incorrect username and password access should be denied.	Access was granted to users that enters correct username and password and also access was denied for users that enters incorrect username and password
3	Users clicks on upload files	Users should be able to upload files for backup	Users was able to upload files for backup
4	Users clicks on message	Users should be able to send messages across to admin	Users was able to send messages across to the admin
5	Admin click on view backup,	Admin should be able to view backup	Admin was able to view backup

The above table illustrates how the software program was tested

Performance Evaluation

Performance evaluation is the process of checking if the system performs as required and satisfies certain conditions. When the expected and actual results were compared, there was no error and the system met the conditions stated in the requirement and specification.

Software Integration

Integration Testing is a level of the software testing process where individual units are combined and tested as a group.

System integration is the bringing together of the component subsystems into one system and ensuring that the subsystems function together as a system. In information technology, systems integration is the process of linking together different computing systems and software applications physically or functionally, to act as a coordinated whole. During system integration, each module was tested before integrated to another module, there was also

testing at each integration. A final test was done when the modules where integrated into the complex system.

5.0 Conclusion

In conclusion, Storage Area Networks (SANs) have emerged as the best solution for advanced storage requirements. The (SAN) answer the information infrastructure and application needs of today's corporations both large and small, government or civilian. SANs provide scalability for cost effective configurations as well as expandability. They provide flexibility for ease of implementation and maintenance. SANs are also reliable and redundant, providing integrity and security for data. Finally, SANs allow clusters of storage devices to reside in different geographic locations for support of offsite data storage. Storage Area Network technologies enjoy wide industry support with industry associations, well tested standards and multiple vendors.

References

[1] http://www.en.wikipedia.org cited on August, 2015

[2] Laura Lee Brennan and Marc Olanie (2007). An Introduction of Network Attached Storage, SPH magazine

[3] Ugah, J.O. et el (2011), A modern approach to computer application, His glory publication, Enugu Nigeria

[4] Matthew Glidden (October 2001). Switches and Hubs. About this particular Macintosh blog. Retrieved September, 2015

[5] Robert J. Kohlhepp (2000) The 10 Most Important Product of the Decade."Network Computing magazine

[6] http://www.redbooks.ibm.com/redbooks/pdfs/sg245470.pdf

[7] http://www.cisco.com

[8] http://compnetworking.about.com/od/networkstorage/g/storage_san.htm

[9] Feig, Rami (2001). Computer Networks: The OSI model Reference model http://www.rad.com/networks/1994/osi/intro.htm