Enterprise Storage Management System

Dan Glasser[1], Madeline Hardojo[1], Anand Sundaram[1], Nate Wells[1]

{dglasser, anands, mhardojo, nwells}@cse.unl.edu
and
Mohamed Fayad, PhD[2]
m.fayad@sjsu.edu

Abstract:

Enterprise Storage Management System is an interactive and user-friendly program that will enable the Lincoln Telephone Company to efficiently manage their storage system. With this system, the Lincoln Telephone Company will be able to monitor the file system in the storage, optimally backup; both manually and automatically, recover their data while maintaining its correctness, and provide statistic information as a reference for the system administrator. The system also provides some maintenance functions such as identifying bad or full disks, assigning and redirecting available resources while keeping the integrity of the system, and compressing the inactive files.

Background:

In a world where information is the key to success, every company must take adequate measures to protect its data. In a typical corporate world, tons of information flows across the network that needs to be processed. This data that is being used for processing needs to be stored in a safe and secure medium that is also easily manageable. With the advent of new storage technology and falling price of the storage, IT managers buy more storage devices for their rapidly increasing corporate data. This process in turn makes managing the growing storage very complex.

Lincoln Telephone Company wants to have a solution for the growth of storage in the company. The architecture of Lincoln Telephone Company is illustrated in Figure 1. The company has a number of servers to accommodate its applications and data. When a client requests some information from the system, the server will check whether the desired data is located in the server or not. If it is not, the server will locate the data from the various backup disks. They would like to have this process automated with a user-friendly system.

Description of Program that is Wanted:

The storage management system should address the following issues:

➢ An efficient mechanism to backup and recover the data – this may include a distributed backup or centralized backup

- ➢ Monitor the file system in order to automatically assign and extend additional available resources
- ➢ Replace a backup resource with an available resource if the backup reaches a threshold capacity.
- ➢ Mechanism to format the available disk with the required File System in a multi platform network
- ➢ Add/Delete resources to/from the storage network in a multi platform scenario
- ➢ Report backup failures due to network failures. Reschedule the incomplete backup tasks as soon as the network is up.
- ➢ Identify bad disks in the network and remove them. Backup/recover the disk before doing so.
- ➢ Automatically identify the resources that are added or deleted to the network.
- ➢ Create an Analysis report of the backups that will help users schedule the backups efficiently.
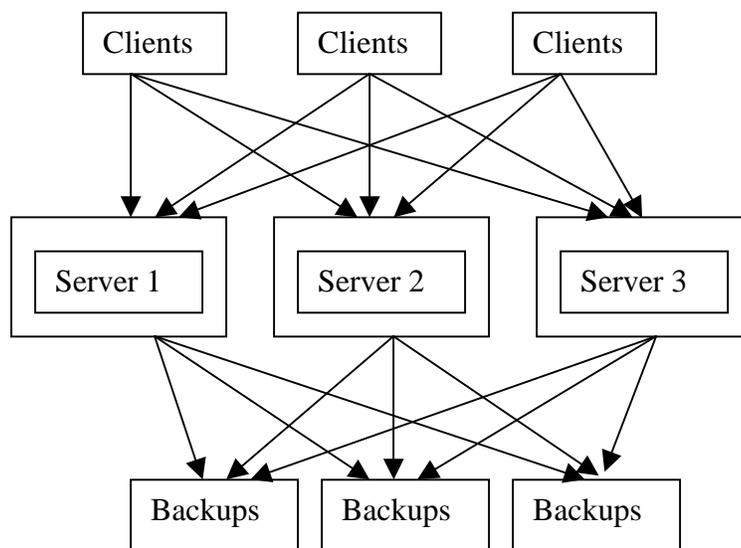- ➢ Identifying dormant files and automatically compress them to be archived in a backup disk

Fig. 1 - The Architecture

Detailed Requirements:

Lincoln Telephone Company, for protection against destruction of its data, either accidentally or intentionally, feels the vital need to back up its valuable data such as customer and billing information on a daily basis; sometimes even hourly basis. Lincoln Telephone Company's backup servers are kept in geographically distributed buildings. The Enterprise Storage Management System should be able to manually and automatically backup the data and be able to recover the data if needed. A system operator should have the ability to set the backup schedule and frequency. Also, the

system should be able to recognize if the data is corrupted and should be able to pull out the latest backup and recover the lost data.

As we can see in Fig. 1, the servers of Lincoln Telephone Company are accommodating applications and data. The storage of data in the servers may be full. Therefore, the Storage Management System should be able to detect if the server is nearing capacity. If this condition is detected, the system should automatically allocate a place in the backup disk for the data and store it. This would provide the necessary free storage space in the server for the application or data.

Also, the Lincoln Telephone Company has purchased many disks for backup when the price for storage dropped. All the backup disks, both non-empty and empty, are connected through a network. The company would like the system to be able to notice if the backup disk currently used to store the data from the server is almost full and to automatically redirect the data to the empty backup disk.

When accessing a disk for backup, the user should be able to specify what file format to store the data in. If the user does not care what format the data is stored in, the Store Management System should choose the optimal and convenient format to store the data. In all cases, the system should be able to accommodate a wide range of storage formats. Additionally, the system should be able to provide data in a given format to a user in the possibly different format that they request.

In a multi platform network, the requirements for the storage media are very high. The system should be able to easily detect any new resources that have been added to the network. The system should also be able to remove any resource from the network that is no longer needed or needs to be relocated. This will help the storage management system to manage the growing requirements of the storage disks by allocating the free available disk for backup purposes. Essentially, whenever a backup disk is full, a new free available storage disk is identified for replacement. This is done by checking whether a free disk is available in the network by Handshake mechanism. Once a free disk is found, it can be formatted with the appropriate file system and be marked as a backup disk in use.

In a network where all the servers, storage resources and other systems are connected, there might be failures due to network problems. In order to address this issue, the storage management system should be able to detect the storage related problems and fix it automatically. If the backup fails because of network failures, the system should be able to detect it. The problem can be reported to the administrator. As soon as the network problem is resolved, the backup process that has failed should be automatically rescheduled.

The data that is stored in the disks (either backup or normal) should be protected. If there are problems like data corruption or hardware problem, the system should be able to resolve it. The system should run a tool that will perform data integrity check on each of the storage disks in the network on a regular basis. If the system finds data corruption, it should be fixed. If there is any hardware issue, the disk should be replaced with a new disk after restoring the data from the corrupt disk.

In order to have an efficient backup mechanism, the system should be able to do the backup without adding much traffic to the network. To achieve this, the system should be able to schedule the backup processes when the network traffic is minimal. The traffic in the network system can be studied and a report of peak load and trough network

usage can be produced that will help the administrator schedule the backup process efficiently. This process can also be automated.

Some of the files that are backed up by the Lincoln Telephone Company are used for a period of time and then become obsolete, and thus their usage ceases. This is an inefficient use of storage space. The Lincoln Telephone company would like a system that will be able to periodically check the whole backup disk system and automatically compress these inactive files in order to optimize storage space. The system would also need to create a log of the files that being compress for the Administration purposes.
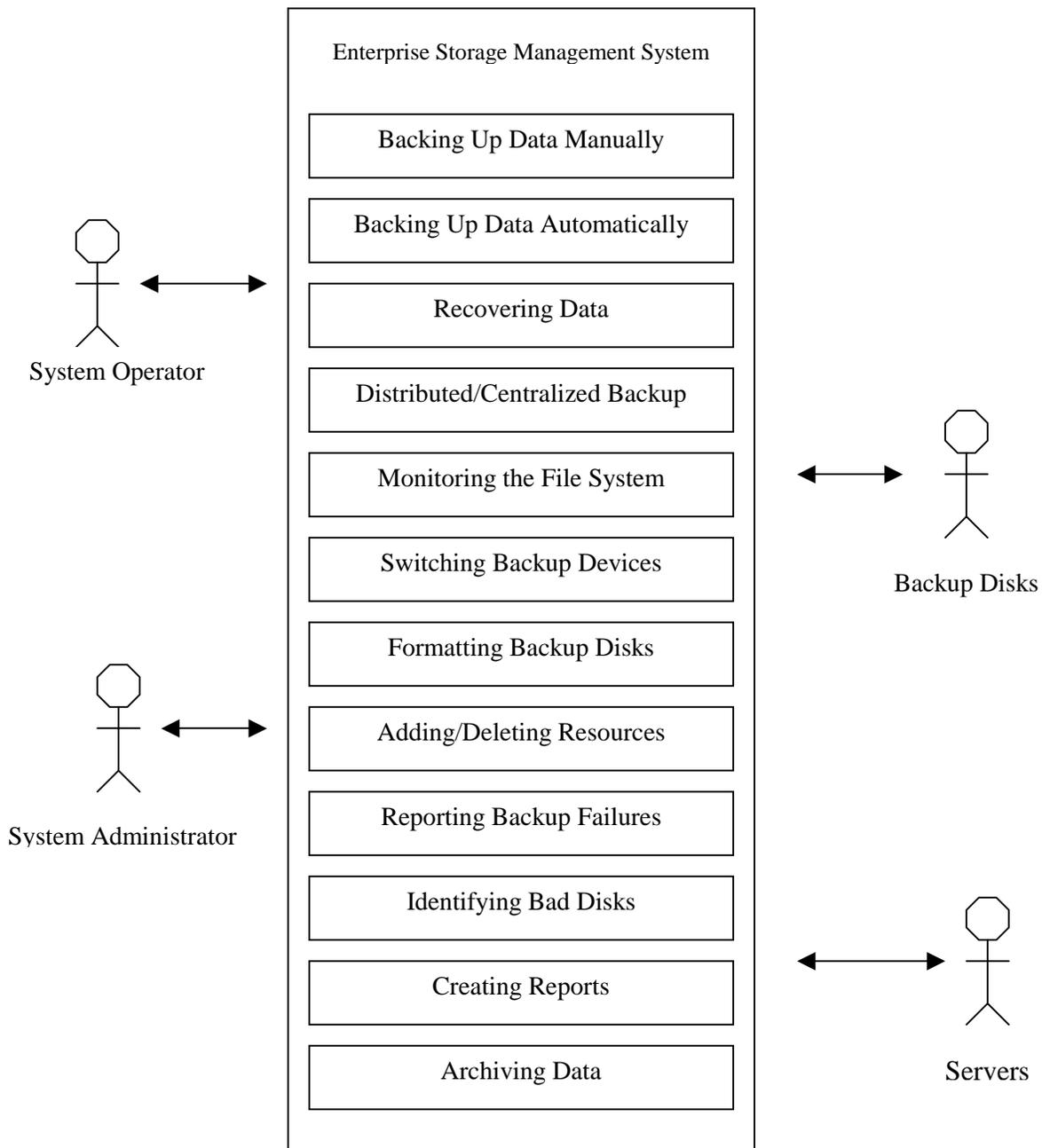
Use Cases:

Case 1: B

Fig. 2 – Use Case Model

The operator for the Enterprise Storage Management System has received a request from the Billing Department to do a backup of all the billing transactions for the current month. This will be done as safety measure since the Department is going to implement a new program to manage billing transactions. For security and for future reference, the operator will then run the Enterprise Storage Management System and enter on his/her operator id and from whom the request was received. He/she will then go to the backup interface and choose which directory or file to be backed up. The system will then automatically allocate a space in the backup disk and name the directory/file appropriately so that it will reflect necessary information, such as the date of backup and the original directory/file name. Additionally, it will generate a "readme" file that will contain information of which operator did the backup and other related information.

Case 2: Backing Up Vital Company Data Automatically

Lincoln Telephone Company is a fast growing company. Therefore, it is necessary to backup their new customer information daily. Of course to do this manually would be such a tedious job, not to mention, it would require an operator to be responsible and to spend his/her valuable time every morning to perform the backup. An operator would then have to enter the backup interface of the system and specify which directory or file to be backed up. Then, he/she will input the needed information to generate backup automatically. This information includes the time of day the backup should occur and how often this should be repeated. This information should be able to be modified at anytime. On the schedule given, the System will automatically backup the directory/file specified. In addition, it will generate a report log of the necessary information, such as which directory/file is backed up, etc. In case of the server failure or any case when the System were to fail to back up the information, the system will automatically e-mail the System Administrator Department to warn that the automatic backup failed and therefore manual backup and other adjustment is needed.

Case 3: Recovering Data

The System Administrator received a call from Financial Department. They are frantically reporting a crash in the system and would like to have their data back since tomorrow is payday. The data in the Financial Department is being backed up every week. Knowing that, the System Administrator later called the operator of the Enterprise Storage Management System so that he/she can recover the data that was lost in the crash. The operator will identify from the system the latest data available in the backup system. The system then will look through the backup disks to find the files/directories needed, retrieve them, and recover as much of the lost data as possible.

Case 4: Distributed / Centralized Backup

Lincoln Telephone Company has a multi platform network in which the storage can be made in a centralized backup disk or distributed backup disks. The system has the functionality that will help the Administrator decide to either distribute the data storage or store them in a single disk based on the requirements of the application. For a centralized backup option, the administrator has to specify the available storage disk. If the disk becomes full, the storage management system detects a free available disk in the network that can be used to replace the full disk. In a distributed storage option, the administrator gets to choose the disks from a set of available disks. The administrator can also automate the process by letting the storage system do the data storage in a distributed environment.

Case 5: Monitoring the File System/Server Storage

With a high storage requirement for Lincoln Telephone Company, it is critical for the company to monitor the growth of the data. The rate of the space utilization by the applications determines the need for additional storage disks. A smooth transition must be made when a disk in production becomes full. The company can do this by defining a threshold value. If the utilization of disk space crosses this value, the storage management system will alert the administrator accordingly. The system will also identify the free available storage disk for extending the storage. Once identified, the disk is formatted with the required file system and is marked as live. Now, the new disk runs in parallel with the disk that has just become full.

Case 6: Switching Backup Devices

There is also a need for the system to switch between the multiple backup disks. The need could arise when moving data to a backup device that is nearly its capacity. The Enterprise Storage Management System will discontinue using that device and start using a different device that has space available. The first device will be marked so that it is no longer available to any additional processes as a place to write data. Additionally, the transition to a new backup storage device will be performed so that both devices are still accessible by any process that requests any of the data.

Case 7: Formatting the Disk

Since the company data is highly critical, it brings forth the need for data backup. Essentially, whenever a disk in use becomes full (crosses a threshold value set by the administrator), it must be replaced with the free available disk in the network. If no free disk is available in the network, the administrator must be notified well in advance and the administrator will be able to arrange to have free disks available in the network. Before replacing the disk that is full with a free available disk, the new free disk must be formatted. The system allows the administrator to format the disk with any of the supported file systems. The system lets the administrator choose the file system that the disk is to be formatted with, from a list of supported file systems. The system also gives a

table consisting of information on the file systems supported by the various operating systems. This information should help the administrator in making the decision to choose the right file system. Once chosen, the disk is formatted with the specified file system and marked "disk in use"

Case 8: Add/Delete Resources *and* Identify Newly Added/Deleted Devices

Whenever changes are made to the physical makeup of the network, the Enterprise Storage Management System will detect the changes and update the configuration of the System automatically. This means that there is no need for an administrator to go through the steps of allocating the new device and reconfiguring the System. For example, if new storage devices are added to the network, the System will detect those devices, format them as needed, and allocate the devices as primary storage devices or backup storage devices. This "plug and play" characteristic of the System gives administrators an easy way to increase their storage capacity, which is very important to the quickly expanding Lincoln Telephone Company.

Case 9: Reporting the Backup Failures

Lincoln Telephone Company realizes the significance of the data backup process. The storage management system is able to provide just that. The system allows the administrator to schedule the backup either manually or automatically. The system is very robust. It makes 100 percent sure that the data is not lost during the backup. The system is also capable of handling situations like network failures, data corruption, hardware failures, etc. In the event of network failure, the system alerts the administrator that the network is down. During the data backup, the system logs the state of the backup process in a system file. If there is any backup process that is running when the network failure occurs, the backup starts from the point where it failed last as soon as the network is up. This is also called incremental backup. This can be done automatically.

Case 10: Identifying Bad Disks in The System

There are some occasions in Lincoln Telephone Company when a disk can be damaged or corrupted. The operators need a mechanism that will allow them to avoid these corruptions. This is done with a Scandisk system. The operator can set up a scheduler that will periodically run the Scandisk. The Scandisk will search the backup disks for bad sectors. When found, the operator will be alerted that the disk is faulty and must be replaced. Meanwhile, the system will redirect the data from the corrupt disk to the new replacement disk. A status report will be generated indicating what portion of the data was salvageable and which files could not be saved.

Case 11: Creating an Analysis Report

Lincoln Telephone Company requires that the backup processes be scheduled such that the network traffic is not affected to a great extent. In order to achieve this, knowledge on the network traffic during a busy day is required. The best method to do

this is to study the network and come up with an analysis report. The storage management system has a tool that exactly does that. The tool collects a statistics of the network traffic and creates a report. This report will help the administrator to schedule the data backups such that the network traffic is not affected. The administrator may decide to plan the schedule when the network traffic is minimized, say on weekends.

Case 12: Archiving Dormant Files

Lincoln Telephone Company creates many files purely for record keeping, such as financial records and summaries. Once a new fiscal year is started, those financial records, while important, are no longer relevant to the company's current activities. Instead of letting these files take up precious storage space, the Enterprise Storage Management System will compress these files after they have not been accessed for given length of time, determined by a system administrator. Thus, the use of the storage space containing those files will be more optimal. More generally, the System will regularly examine the files it has stored for any files that have not been accessed for that length of time. If it finds any dormant files, it will compress these files, noting it in a log file. The files will still be available upon request, however they would be stored in a compressed format, allowing for more efficient use of the storage space.

Interfaces:

The Enterprise Management Storage System will be required to interface with all other applications that Lincoln Telephone Company may be using, in the sense that it will need to protect and manage the data used by all other applications. However, there is no need for a direct interface between the System and other applications because the System does not directly change the data, but rather just back it up or possibly move it to a different storage device. In that sense, it is insignificant which applications the Lincoln Telephone Company uses, what format those applications use to store their data, or even what operating system those applications are run in. The System will protect any data created by any application in an efficient manner.

References:

Tivoli® Storage Products by IBM. http://www.tivoli.com