A Case for Automated Tape in Clinical Imaging

Geoffrey Bookman and Donald Baune

Electronic archiving of radiology images over many years will require many terabytes of storage with a need for rapid retrieval of these images. As more large PACS installations are installed and implemented, a data crisis occurs. The ability to store this large amount of data using the traditional method of optical jukeboxes or online disk alone becomes an unworkable solution. The amount of floor space, number of optical jukeboxes, and off-line shelf storage required to store the images becomes unmanageable. With the recent advances in tape and tape drives, the use of tape for long term storage of PACS data has become the preferred alternative. A PACS system consisting of a centrally managed system of RAID disk, software and at the heart of the system, tape, presents a solution that for the first time solves the problems of multi-modality high end PACS, non-DICOM image, electronic medical record and ADT data storage. This paper will examine the installation of the University of Utah, Department of Radiology PACS system and the integration of automated tape archive. The tape archive is also capable of storing data other than traditional PACS data. The implementation of an automated data archive to serve the many other needs of a large hospital will also be discussed. This will include the integration of a filmless cardiology department and the backup/archival needs of a traditional MIS department. The need for high bandwidth to tape with a large RAID cache will be examined and how with an interface to a RIS pre-fetch engine, tape can be a superior solution to optical platters or other archival solutions. The data management software will be discussed in detail. The performance and cost of RAID disk cache and automated tape compared to a solution that includes optical will be examined.

Copyright © 1998 by W.B. Saunders Company

RADIOLOGY DEPARTMENTS over the last few years have begun to change dramatically. The traditions and practices of radiology departments in creating, handling, and storing films are giving way to computer aided methods of providing better patient care. PACS systems and other electronic image management tools are begin-

ning to enjoy expanded use in radiology departments. The major impediment in the wide use of electronic radiology imaging continues to be the difficulty of archiving and managing images once they have been used in the diagnostic process. PACS systems traditionally have not saved radiology departments money because film is still being produced rather than being stored electronically.

Today automated tape holds the key to better patient care in radiology. Its higher capacity, high performance, ease of management, robust and cost effective qualities make it the technology for today and the future. Tape has suffered in the past by having a reputation for being less than robust. Tape today no longer stretches, or needs to be cleaned. It has a life of 30 years (rivaling optical) with a magnetic loss of less than five percent. Coupling automated tape qualities with robotics is an elegant solution to today's radiology department image management problems.

CENTER FOR ADVANCED MEDICAL TECHNOLOGIES CONFIGURATION (CAMT)

The configuration at the medical center consists of hardware, software, and the modalities being managed by the system. The CAMT had a need to store and manage DICOM and non-DICOM data within a central archive system, that could be easily managed from a single workstation. Because of the critical nature of the data, a Sun Enterprise 5000 server was chosen for control of the system. It is a powerful processor with many redundant components, including hot swappable CPU's, motherboards, I/O cards, and disk, allowing the system to remain in operation even when individual components fail. It is important to any radiology department that patient data always be accessible.

A StorageTek (Louisville, Colorado) Clariion RAID5 disk subsystem with 430 gigabytes of storage was chosen to serve as cache for all data including DICOM images and general data. The patient data is stored on the RAID5 disk for complete safety and data integrity. This data is kept on RAID disk until certain, definable parameters are met, and then, using HSM software, swept off to StorageTek 9710 robotics library with DLT7000 tape drives. Each modality or data type can be managed automatically by its own characteristics,

From the Storage Technology Corp, Salt Lake City, UT; and the Department of Radiology, Center for Advanced Medical Technologies, University of Utah Medical Center, Salt Lake City, UT.

Address reprint requests to Geoffrey Bookman, Medical Applications Consultant, Applications Business Group, Storage Technology Corporation, 349 South 200 East Suite 540, Salt Lake City, Utah 84111.

*Copyright © 1998 by W.B. Saunders Company
0897-1889/98/1103-1014\$8.00/0*

once the policies for that data have been determined and set.

A StorageTek Timberwolf 9710 robotics library was chosen as the long term archive for the data because of its robust qualities, speed, and ease of use. This model will hold 20.5 terabytes of data (uncompressed) and can contain up to 10ea. DLT7000 tape drives for performance, redundancy, and data throughput needs. The robotics library holds up to 588 tapes with a capacity of 35 gigabytes per tape: 20.5 (Terabytes uncompressed) for the library. This is equivalent to about 4-5 years of patient data for the University of Utah Medical Center (385 beds). This library is designed to operate at 180 tape exchanges per hour (retrieve and storage of tape) 365 days per year and 24 hours per day.

Perhaps the most important aspect of this configuration is the data management. Veritas (Mountain View, California) software was selected because of its functionality and ability to easily manage large quantities of data. Veritas software has the ability to manage all the radiology images as well as Cardiology, Genetics, and others. Veritas software will also enable the hospital to backup the entire enterprise from anywhere in the hospital sphere of control. This ability was the defining factor in choosing Veritas for data and media control. The components are Media Manager, Veritas NetBackup, and Veritas HSM.

IMAGE WORKFLOW CONSIDERATIONS

In Filmless Radiology and Cardiology, correct size RAID, pre-fetch, and auto routing is the heart of productivity: getting the image to the right place on time. In a traditional radiology department a small army of file-room clerks will pull the films from the short and long term archives. These images are then placed in the appropriate reading rooms. After an exam is completed it and the selected priors are placed on the reading rooms viewing boards. The way in which films are placed on the reading boards is very specific to the radiologists and the type of exam that was ordered. Historically these are some of the areas in which traditional PACS systems are weakest. Then after a study has been read the entire folder of images must be sorted, placed in their film jacket and then stored.

When a department makes the transition to filmless readings these same functions have to be

accomplished. In the PACS system that we are implementing the Radiology Information System (RIS) controls most of these functions. The RIS that we utilize is IDXrad (Burlington, VT). The evening before the scheduled exams, IDXrad issues a series of requests (known as a pre-fetch list) to the image archive to retrieve all of the images of the patients that are scheduled to be seen the next day. As an exam is completed on a scanner (MR, CT, CR, Nuclear, Angio, and/or Ultrasound) the images are automatically stored in the DICOM archive. Then the auto routing engine looks up in its rules what prior images belong to the exam that has just been completed. It also looks at where this type of study is read and then the appropriate images are moved to the correct viewing station. Future work will include the ability to enter which physician is reading at an alternate viewing location enabling the movement of images based on a work schedule instead of a static routing list. Once the images are located on the viewing station the physician then selects all unread cases and works through the daily workload. As a case is dictated the RIS is notified by the dictation system that the case has been dictated.

With the images all residing on a central archive any physician can, from any workstation, request that a patient's 'jacket' be sent to their view station for review. This can happen when they are researching a specific topic, a patient is calling about a future appointment, or another physician is calling for a consultation. The RAID is sized to hold about 3 weeks worth of patients. This includes the prior and the current patient study. After this time period if there is no further activity on this patient, the entire 'jacket' is then re-archived on the automated tape system. After the images are written back out to the tapes the previous images on another tape are nullified. While this may seem inefficient it keeps a patient's entire directory of images together so that only a single seek is necessary to retrieve an entire patients history. As tapes become fragmented they are automatically compacted by the Media Manager that will be described later on. The ability to rewrite and reuse tapes that have been previously used is one feature of tape that is superior to optical disks. While there are re-writable optical disks, they have not historically been used in PACS systems.

A second major advantage of automated tape is

the ability to keep a patient's entire history of images on a single tape. Optical disks have the ability to do random seeks for data which is necessary as the patient's data is often distributed over several disks or in many locations of a single disk. The transfer time of tape for a relatively large amount of data is much faster than is available from optical disks. We have experienced a maximum unload of the previous tape, load of correct tape and the retrieval of an entire patient image directory in an average of 84 seconds. When the patient record is at the beginning of the tape, retrieval takes about 24 seconds.

DATA MANAGEMENT SOFTWARE

The Veritas software was selected as the Hierarchical Storage Management (HSM) system. This software package was selected because of its ability to handle the many variations of our total data archiving and backup needs. There are many software packages available to move data in an HSM fashion or to do network backups, however there are very few that can do both and share a single archive in the process.

Veritas is also one of the few data management software companies where the entire software package is written in-house. While it is not necessarily bad to subcontract pieces, it can slow down the product development and testing process. Also if a problem is discovered in the field it is easier to debug and fix if the entire software package belongs to one company.

The Veritas software is made up of three major components which when combined make up the storage management system. The components are a Media Manager, Veritas Netbackup and Veritas HSM. The Media Manager controls the robotics and the automated tape movement requests. It also has a complete inventory of the tapes in the archive and of the tapes that have been removed from the archive for disaster recovery. It is the job of the Media Manager to manage the database of tapes and their contents for Veritas NetBackup and HSM components as requested. When new tapes are added to the archive they are added to either a Netbackup pool or an HSM pool. The Netbackup component is made up of a server and client pieces. The server runs on the same computer as the Media Manager. The client pieces are installed on various computers in the network that have backup requirements. When the client runs the backup software it

is sent across the net to the server component of the Netbackup application. Users can initiate the back-up requests or they can be automated to occur at any time of day. The user or the system administrator can initiate requests for a recovery of a file that has been backed-up to automated tape. This ability for the user to retrieve a lost file can save many hours of delay from the time that a file has been lost and the time it takes a system administrator to restore it. This will also reduce the workload on system administrators.

The HSM component hides the fact that data has been migrated to automated tape by making it transparent from the end user. When a request for archived data is made, the HSM retrieves the data from tape and replaces it on the RAID. The only thing that the user sees is a slight delay in the response time to his/her request. The main advantage to the user is the unlimited supply of disk space (This has been compared to having the Holy Grail). As discussed earlier, it is critical that the RAID be sized sufficiently so that the majority of the data needs can be stored on magnetic disks and that requests to the archived data is truly a rare event. The 30-day rule is currently in place for the Department of Radiology at the University of Utah. If the data has not been accessed for 30 days it then becomes a candidate for archiving. The current RAID cache system can double in size within the existing cabinet, using 9 GIG disk drives. As new technology becomes available and stable (18 or 27 GIG disk drives), the system will continue to grow with the needs of the CAMT with little or no interruption.

So what does all this have to do with the DICOM database? Several vendors of PACS system data archives have gone to great pains to customize their interface to a specific HSM software package. In the process of doing so they have made the data repository a proprietary resource to the PACS system and therefore cannot be used for the MANY other data needs of a typical department. At the University of Utah this just was not an acceptable solution. We searched for a software only solution to the PACS archive. In so doing we were able to specify which HSM we were going to use. Other PACS vendors will contend (and probably prove) that this approach is not as efficient as a proprietary system. While this may be true, the performance that we are able to achieve from this configuration has been superior for our department PACS needs. Future

research and system use are needed to see if this will hold true as several years of data is archived in the system. However, we have every reason to believe that there is sufficient spare capacity in our archive to meet the needs of a busy department PACs and the data needs of the administration functions, RIS, research data, non DICOM clinical data and the Cardiology Department DICOM image storage needs.

PERFORMANCE AND COMPARATIVE COST

Relative performance, safety, and cost per megabyte stored are important factors in choosing the archive system. Tape today is safe for all patient data requirements. Some would argue that data on optical disk has greater life than tape. Most experts agree that the longest time optical disk can be considered safe is about 30 years. This is just about the same life as DLT7000 tapes. More important to explore is why data should remain on the same media for thirty years. No reading device today will be viable in even a few years. It is mandatory that data be moved to new media as new technology becomes available. This will occur every few years. It is important that each medical facility be provided with an upgrade path that does not greatly impact the operation of the system, and is not prohibitively expensive.

Performance was another factor in choosing

DLT automated tape as the long time archive media. Today, DLT7000 tape can transfer data at a rate of 5 megabytes per second uncompressed. This is at least twice the speed of the fastest optical system and with 2:1 compression, the data transfer rate is five times the transfer rate of optical.

The ability of the robot was also important in making this decision. The STK 9710 library can withstand thousands of automated tape exchanges without failure. The Mean Time Before Failure (MTBF) is 70,000 hours and Mean Exchanges Between Failures (MEBF) is 1,000,000 while Mean Time To Repair (MTTR) is only 30 minutes.

The last consideration is cost per megabyte stored. DLT7000 tapes hold 35 GIG per tape (uncompressed) at a cost of about \$75.00 per tape. This is about \$0.00214 per megabyte. Compare that to Optical with 2.6 GIG per platter and \$75.00 per platter and the cost difference is significant. Magneto Optical is \$0.0288 per megabyte: about 10 times more expensive. In addition, the library cost and capacity need to be taken into account, and this factor adds greatly to cost as well as ease of use. The goal must be to store and retrieve data without human intervention and with total confidence that the patient data will be there. This has been achieved at the University of Utah Center for Advanced Medical Technologies.