SIIM SOCIETY FOR IMAGING INFORMATICS IN MEDICINE
Innovating Imaging Informatics

Archiving, Chapter 5: Data Storage Management

SIIM Archiving: Fundamentals of Storage Technology

Author:
Robert Glicksman, Medical Imaging Technical Specialist (Retired)

Introduction

Medical image archives can grow to contain enormous amounts of data. Radiology, Cardiology, or Pathology image sets can consume several gigabytes of storage. The sum total of image data associated with many patients and over many years can run into the petabyte (10^{15}bytes) range. Obviously, a practical medical image archive must economize on the cost of raw storage.

Many technologies exist to store digital data. Generally speaking, the more accessible (faster access) the data, the more expensive is the storage. Consequently, medical image archives usually consist of a mixture of storage technologies - fast, expensive storage for images that are deemed to be "in use" or "needed shortly" and slow, inexpensive storage for images that are deemed to have a low probability of clinical use but must be retained for other reasons (e.g., medico-legal or medical research purposes).

Here is a summary of key technologies in cost/performance order:

- Solid state storage (RAM, flash memory) [highest cost, very short access time]
- Spinning magnetic disk storage (Winchester disk)
- Optical disk (MOD, writable DVD)
- LTO tape, in a robotic tape library
- LTO tape, externalized from the library ("shelf storage")

Healthcare is episodic. Image data tends to move between categories of "in use," "needed shortly," or "archive purposes only," based upon clinical events. In order to optimize cost and performance, medical image archives generally arrange their mixture of storage technologies in hierarchical tiers. The top tier consists of the fastest and most expensive storage. Lower tiers have progressively lower cost and longer retrieval times. Various triggers may be used by archive management software to anticipate data availability needs and corresponding data migration needs between tiers of storage. Common triggers include:

- The age of the image (the older the image, the lower the anticipated need to retrieve it)
- Patient admitted to the emergency room (possible acute care need related to a recent or chronic illness)
- Patient admitted to the hospital (possible acute care need related to a recent or chronic illness)
- Patient scheduled for an inpatient or outpatient visit or procedure (possible follow-up to an ongoing treatment or chronic condition)
- Patient arrives for a new procedure (possible follow-up to an ongoing treatment or chronic condition)
- Court order for discovery and disclosure (medico-legal)
- Commencement of an epidemiological study (medical research)

These trigger events may be processed automatically by the archive management software or they may be manually processed by an archival administrator who performs administrative actions to move data between tiers in the archive.

In addition to the aforementioned events that trigger data migration between tiers of storage, there are technical events that require data to be migrated within a medical image archive. Technology obsolescence and equipment failure/replacement are common reasons for replacing storage devices with new ones, necessitating the migration of data from the old device (or a redundant copy of the data stored on the old device) to a new device or technology.

Clinical Image Lifecycle Considerations

Medical images are created mainly to support clinical diagnosis (Dx) and/or medical treatment (Tx), and only occasionally for purely research purposes. The episodic nature of healthcare means that both Dx and Tx images have their greatest value at the time of treatment of the patient, which is often a span of a few days to several weeks. Images associated with the Dx or Tx of chronic diseases may need to be retained for a number of years to support ongoing treatment. These older images are generally used for comparison and trending purposes. After a period of 6 months to 2 years, there is less than a 5% chance that an image will ever be requested from the archive.

Generally speaking, medical images must be retained in an archive for longer periods of time to meet legal/evidentiary requirements. The time period depends both on the nature of the disease and the local legal standards or precedents. In the USA and EU countries, images may be required to be retained for legal purposes for a period of 5 to 10 years. However, images associated with certain occupational illnesses may need to be retained for an even longer period of time; sometimes for the patient's lifetime.

In addition to legal retention requirements, medical images may be a valuable source of information for medical/epidemiological research. As such, it may be important to retain these images for very long periods of time (decades).

When an image has passed completely through its lifecycle, it may be deleted. Some healthcare providers believe that it is important to delete such images immediately at such a point in time, and to do so in a manner that guarantees that the images cannot be reconstructed, by any means. Since we are talking about archiving images in electronic form, it is not sufficient to simply "delete" the file entry in a file system, as the image data may still be retained in storage and recoverable via an administrative or service tool. This is particularly true of redundant storage strategies such as RAID and CAS. In addition, all entries in the archival database that are associated with the image may need to be deleted as well, as the metadata alone may have evidentiary value.

Image deletion with assured data destruction is not a simple matter. Consider, for example, image data stored on tape. Images that are eligible for deletion may coexist on a volume of tape with images that must be retained. It may, therefore, be necessary to create a new volume of tape with only the images to be retained and to completely erase (e.g., overwrite or physically destroy) the old tape volume, after verification of the rewriting. It must also be assured that the archival database has the pointers to all new locations for the images, and has deleted locations where the image no longer exists (note: the whole DICOM hierarchy in the archival database of the image may need to be deleted if the last copy of the image is so deleted). In doing so, the status of database backups must also be considered, as an older backup may still contain legacy metadata and obsolete image location information. The destruction of images on non-erasable optical media can also be a problem; as such media may need to be physically destroyed after rewriting images that must be retained onto new media volumes.

Matching Cost/Performance of Storage to the Medical Image Lifecycle

The discussion of clinical image lifecycle above, clearly suggests that a medical image archive could benefit from a hierarchy of storage technologies. Fast, magnetic disk storage might be used for the archival database and for storing images for a period of 6 months to one year. After this period of time, the low probability of retrieval, as well as the more predictable nature of need, may allow the image to reside on lower cost magnetic tape. Many vendors of medical image archives offer the capability to configure the archive from a multitude of storage technologies and for the archive administrator to define rule sets for the migration of image data across the tiers. Figure 5-1 illustrates a typical archival storage migration strategy that would be implemented by such a set of rules.

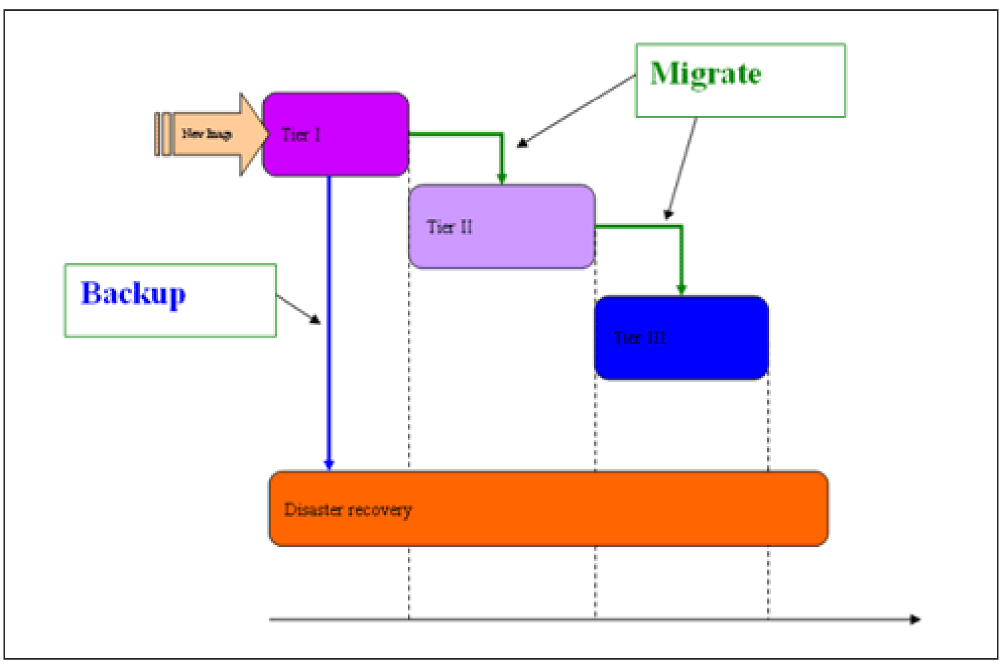

Figure 5-1: Archival Migration Across Storage Tiers
(Courtesy of Neville Skudowitz, Carestream Health, Inc.)

Referring to Figure 5-1, the tiers of storage may be as follows:

- Tier 1: Fibre-Channel RAID
- Tier 2: SATA RAID
- Tier 3: Tape media, in a tape library

- Disaster Recovery Tier: Tape media, externalized from a tape library ("shelf storage") or redundant RAID located in another geographic location

The archival storage management rule set would be programmed by the archive administrator, according to Figure 5-1:

- New images are stored on Tier 1 for a period of 6 months
- New images are stored on tape media; tape volumes are externalized and stored offsite daily
- After 6 months, images are stored on Tier 2 storage and, after confirmation of storage on Tier 2, the Tier 1 copy is deleted
- After 2 years, images images are stored on Tier 3 storage and, after confirmation of storage on Tier 3, the Tier 2 copy is deleted
- After 7 years, the Tier 3 copy is deleted
- After 10 years, all remaining copies of the image are deleted and the database entries that relate to the image are also deleted. Note: this step may require mounting externalized tape volumes and re-creating partial tape volume content, as noted above

The clinical image lifecycle rule set may need to be much more complex than the simple rules illustrated in Figure 5-1. For example, the lifecycle management rules may need to be set based upon exam/procedure type, imaging modality, disease, research tags, and other types of information. Many vendors of medical image archives support sophisticated rule set generation. However, it should be noted that the necessary metadata may not be available to implement a very sophisticated set of rules. In general, exam/procedure type and imaging modality can be expected to be part of the DICOM metadata; disease and/or research tags may need to come from other IT systems.

The topic of Hierarchical Storage Management (HSM) software was briefly introduced in the previous chapter. Commercial HSM packages (e.g., IBM's Tivoli) can manage the migration of data across multiple tiers of storage based upon general purpose rules, such as "least recently accessed" or "least recently modified." These general purpose rules do not fit well with the clinical image lifecycle. Vendors of medical image archives usually create their own image management middleware that is clinically aware and can use information encoded in DICOM and HL7 tags. Nevertheless, HSM software may still have use in the archive as a means to provide the vendor's middleware with file level access to image data, regardless of the underlying storage technologies used.

Pre-fetch and Pre-load

The clinical image lifecycle may be exploited to optimize cost and performance within a medical image archive, as described above. However, medical images are rarely viewed directly from the archive. Rather, medical images tend to be viewed within the context of a clinical information system, such as a radiology PACS, cardiology PACS, pathology reporting system, etc. These external (to the archive) systems may simply request the images back from the archive; usually via DICOM query/retrieve services. However, many such systems are not aware of the archive and, even if they are archive-aware, the images may be required well in advance of the viewing activity (e.g., for pre-staging, "hanging", of images in some pre-defined sequence). Various clinical triggers may be used to retrieve image from the medical image archive so that they are pre-staged for viewing on the information system well in advance of the viewing need. Examples of such triggers include:

- Patient arrival for a pre-scheduled new exam or procedure
- Ordering of a new procedure or exam of a similar type to prior exams for the patient whose images are archived
- Patient appointment for a follow-up physician visit relating to treatment for, or during which, images were created

- Patient appointment for a consultation with a specialist

Triggers may be used by the upstream information system to "pull" images from the archive, or the archive may have the capability to receive and process such triggers to "push" the images to an appropriate upstream information system. When an external system pulls the images based upon a trigger, the term "pre-fetch" is often used. When the archive itself uses triggers to gather the images and push them upstream, the term "pre-load" is often used.

When triggers are used, whether for pre-fetching or pre-loading of archived images, the performance requirements for the archive are reduced. Pre-fetch and pre-load operations obviate the need for the medical image archive to be highly responsive, as the triggers cause the images to be retrieved well in advance of viewing need. It needs to be noted, however, that such triggers are not 100% reliable and some degree of ad-hoc retrieval of images from the archive must be anticipated.

Technology Obsolescence and Data Migration

We live in an age of rapid technological advances. This is particularly true of electronic data storage devices, where new technologies are constantly being introduced to the marketplace and where older technologies are evolving rapidly. Technological advances are beneficial, in that they bring higher storage capacities and higher performance at a lower cost. Technological advances are also a curse, as they bring rapid obsolescence to existing archives. Some technologies (e.g., WORM) have failed in the marketplace; devices based upon these technologies become unsupportable. Other technologies (e.g., Winchester disk) evolve at such a high rate that a failed device may need to be replaced by a much higher capacity unit. Whichever the consequence of technological advances, the requirement for decades-long storage of medical image data means that technological obsolescence must be addressed.

In the past, dedicated image management systems, such as radiology PACS, were provided with technology-specific archival storage systems. Whenever the archival technology used became obsolete (e.g., MODs), the entire system needed to be replaced, and the data migrated off of the old system before the old system could no longer be supported. This was a huge burden on the departments that owned these systems. The system replacement cost was high, and the cost of data migration (often via DICOM), personnel training on the new system, and other factors added significantly to this already high cost. As a consequence, many IT departments began to think about medical image archiving, not as a departmental role associated with a departmental information and workflow management system, but as a centralized resource that could consolidate the archiving function across many clinical departments. Proper design allows an electronic archive to be architected to deal with technology obsolescence in a more practical and end-user transparent manner.

The characteristics of a medical image archive that supports a variety of technologies in a hierarchical storage arrangement, and with administrator-driven rule sets to manage the migration of data across tiers of the archive, also provide intrinsic support for technology migration. A new technology may be added to the archive with rule sets created (by the administrator) that gradually migrates image data off of older technology tiers and onto newer technology tiers — without the end users (and end user information systems) even being aware that the migration is taking place. Unlike DICOM data migration from an old PACS to a new PACS, only the image data files may need to be migrated from old to new storage tiers; the archival database is maintained intact and is updated reliably to track the movement of image data. The image metadata in the database need not be changed; only the pointers to the image data files. Image data may thus be migrated across tiers of storage "in bulk," without the need for time-consuming retrieval of individual images in order to migrate the data. For example, a whole volume of tape-resident data may be migrated to a new technology, in the order in which data is stored on the tape. This process is many times faster than the process of retrieving each image via DICOM, which might require a rewind and search through the tape for each and every image to be migrated to a new storage device. This built-in data migration feature greatly reduces the overhead associated with data migration, with the consequence of higher performance and smaller degradation in archival performance to the end user while data migration is in process.

Medical Image Archive Administration

Electronic medical image archives provide many benefits to end users and to healthcare provider organizations. Data is stored reliably, can be retrieved based upon many different indexes and queries, can provide very low cost per image, and can service many geographically dispersed user locations. However, a modern, multi-tier, multiple technology image archive can provide many challenges to the persons who have to administer it. It is very important that a medical image archive be provided with the necessary tools for effective and efficient administration. Some of the challenges that a good medical image archive administrative toolset must address are:

- Data storage and database health monitoring
- Security: user account management, access control administration
- Migration/lifecycle rule sets: monitoring, editing, administration
- Network and storage device health and performance
- Planning tools for upgrade and expansion

The first requirement for a good medical image archive administrative toolset is to provide administrative personnel with an immediate snapshot of the overall health of the archive, and proactive notification of new or impending failures. A good way to do this is with a software "dashboard." The concept here is similar to an automobile dashboard — the latter gives the driver an instantaneous presentation of the health and performance of the automobile while driving, without cluttering up the display with too much information (which could actually be negative by unnecessarily distracting the driver). Some key attributes of a medical image archive administrative dashboard should include:

- "Up/down" status for all servers, networks, databases, storage devices and key software processes
- Performance bottlenecks - traffic analysis summary and congestion alarms
- Storage and resource fill - "read/yellow/green" for critical resources
- Device and process failure alarms

A dashboard alone is not sufficient for administration of a complex medical image archive. The dashboard only provides a high level summary of performance and operation, and alarms for critical issues that require immediate administrator attention. A good medical image archive will have a complete set of security and administration tools that provide, as a minimum, the following functions:

- User account and access control administration
- Monitoring and changing lifecycle and migration rules
- Tools to take key resources offline for service while maintaining overall availability of the archive
- Tools to recover data from loss due to device or network failures
- Database administration tools, including:
  - Database backup and recovery
  - Database indexing for performance
  - Database compression, when the size (inevitably) gets too large
- Tools to update/validate database links (pointers) to image data files/locations
- Storage device administration, including:
  - HSM and ASM
  - Library administration

  - RAID configuration
  - SAN partitioning
  - NAS administration
  - Versioning and immutability on CAS and non-CAS storage

- Time-based report generation on:
  - Archival resource utilization
  - Network loading and performance
  - Data access statistics
    - By device
    - By location
    - By user/host system

  - Data access reporting by user/host system (HIPAA compliance)

In addition, administrators should have tools to help them plan for the inevitable expansion of the archives, and for technology obsolescence. Planning tools should include:

- Tools to model archival growth and predict points at which additional device resources will be required for:
  - Storage devices
  - Network devices
  - Server/host devices

- Tools to simulate and model device/system/network performance, as medical image archives can become very complex and it is not always obvious where bottlenecks and chokepoints may appear

In summary, the availability and provisioning of good administrative tools may be the single most important factor in the success of the medical image archive. These tools must ensure that that the archive will:

- Always be available for data storage and retrieval (24x7x365)
- Maintain adequate performance through upgrades, failures and repairs
- Be available to, and accessible by, all valid users, at all times

Tweets

Don Dennison Retweeted

SIIM
@SIIM_Tweets

Exhibit discount for SIIM Corp Mbrs share your #enterpriseimaging solutions at

Embed View on Twitter

Connect