Metrics to Use on the Road to HSM

Jim Gast

Palindrome Corp

Abstract

While the mechanisms of Hierarchical Storage Management may be well understood, the metrics used to measure the effectiveness of HSM and the variables used to justify purchasing an HSM system are not was well defined. Moreover, there are some potholes on the road to configuring an efficient HSM system. This paper suggests a set of steps from justifying the expense to making sure HSM integrates well with the other storage management functions already in place.

There are some practical aspects to converting an organization from voracious disk space consumers into space-conscious informed consumers of hierarchical storage. The steps you take when investigating HSM will determine whether your endusers and applications will be able to survive in a world where old files can be slow. This paper will help you determine an appropriate size for an HSM system for your organization and show you how to make sure other storage managers (like virus checkers and full text indexers) aren't destroying the clues HSM uses to decide which files are least recently used.

Very old files that have not been accessed in many months are moved to optical disk or tape at a fraction of the cost of on-line disks. A phantom file is left behind so that endusers are unaware that the file is gone. HSM systems are designed to automatically de-migrate (restore) the data to on-line disk if the file is accessed. To achieve great economies for large nearline libraries, the HSM hardware is often an autoloader or jukebox.

This grooming process (identifying and migrating files that can afford to be slow) makes the disk appear to be as large as the capacity of the entire autoloader. Level 2 HSM systems[1] make the disk appear infinitely large because they permit multiple tiers of storage and do not require migrated data to stay robotically available.

HSM promises to automate the disk grooming process to avoid the Out-Of-Space disasters. But does HSM itself still need to be monitored, managed, and administered? This paper will introduce you to the tools and metrics of HSM.

The press has helped publicize the experiences of early adopters of HSM. InfoWorld noted "Analysts and LAN administrators working with HSM caution, however, that organizations shouldn't blindly jump on the HSM bandwagon. A close evaluation of your file usage and storage requirements will reveal whether HSM can subdue your storage monsters."[2]

PHASE 1: Measuring WorkingSet

Before you even start moving data, you can determine whether you already have applications in your storage management suite that are bad neighbors.

The WorkingSet is the set of files that are actively being accessed often enough to justify keeping them on fast on-line disk. Notice that "accessed" does not necessarily imply "modified". Most modern operating systems maintain these as independent dates.

HSM systems typically depend on Access Dates to determine which files to delete when space gets tight. If a virus checker steps on all the access dates of the programs every night, the HSM system will think all of those files were accessed last night -- and none of them will be old enough to allow migration. Many backup programs are unaware of access dates and step on them without even realizing it. There are also a few text searching and text indexing applications that have the same un-neighborly behavior.

Look at the access dates of the files in an area of the disk you know has no user access. If the access dates are all set to the date of the last full backup you have a strong hint who the offender is. On Novell NetWare, use the

[1] Graziano, Claudia., LAN Times, June 13, 1994, McGraw-Hill, Inc. Ms. Graziano refers to industry analyst Michael Peterson's identification of 5 separate levels of Hierarchical Storage Management.

[2] "Running out of Room?", Willett, Shawn. InfoWorld, August 1, 1994, InfoWorld Publishing Company.

1063-6390/95 $4.00 © 1995 IEEE

"`/dates`" parameter of the "`ndir`" command. In OS/2 HPFS, Microsoft Windows NT NTFS or Windows '95 FAT look at the file properties. In UNIX, use "`-atime`" on the "`find`" command.

Even applications that don't hurt the access date can cause thrashing, unintentionally recalling large numbers of files from the nearline storage. Some file viewers will "help" you out by showing you a sample of the contents of a file if you accidentally leave your cursor on the filename too long.

Six Week Benchmark

If you have no bad neighbors stepping on access dates, you can proceed. Otherwise, you will have to wait long enough to have access dates that are at least old enough to calculate your WorkingSet size. For this paper, we will assume that a useful approximation for WorkingSet size is the total space consumed by files that have been accessed in the last 6 weeks.

The interesting number here is dormant space (space taken up by files that are not in the WorkingSet). This is space that could potentially be used as free space on an as-needed basis. Be sure to adjust for compression systems whose activity may also (independently) help maximize free space. (For calculations, assume that compression averages 2:1 or use the statistics that come from your compression system. In file-by-file compression systems, assume that all dormant files are compressed. Recently modified or accessed files might not be compressed. You may need to measure or estimate the effect compression has on your WorkingSet.)

Also calculate the size of the AnnualSet, files that have been accessed in the last 52 weeks. This number will be used in the calculation to size the nearline storage.

Measure Again in Six Weeks

TotalGrowth is simply the free space at time t(6weeks) minus the free space at time t(0). During this time some files became dormant and other files were simply modified or accessed.

Typically, the WorkingSet will have grown much less than the TotalGrowth. To measure your likely benefit from HSM, we will need to predict how fast your WorkingSet is growing. These two data points are a crude attempt to estimate that. It will be hard to collect history before you install an HSM system, but you may have to estimate this growth to justify the purchase.

All of your files will logically appear to be available, regardless of their physical placement. The total of all of your data is the FullSet.

Without HSM, you would need enough disk to hold the FullSet. If you do not use compression, your compression ratio is 1. Calculate a conservative approximation for your on-line disk space needs (without HSM) in t=1, 2, and 3 years from now:

Disk(t) = WorkingSet(t)+(Dormant(t)/Compression)

With HSM, the compression benefit will not be as necessary, as the dormant files will all be migrated. The approximation becomes:

Disk(t) = WorkingSet(t)

Each of those disk space predictions should then be translated to financial cost. Michael Peterson wrote an excellent paper[3] for help in approximating the costs of labor to install, manage, back up and maintain the disks. Purchase price of the disk drives is small compared to the labor costs. Finally, don't forget that one WorkingSet megabyte will be 2 physical megabytes if you are duplexing or mirroring. Peripheral Strategies Inc. estimates the total cost for on-line disks to be $16 to $20 per WorkingSet megabyte.

As a rule of thumb, if WorkingSet growth is less than 30 percent of TotalGrowth, the benefits from HSM will be positive. By predicting your WorkingSet growth 1, 2 and 3 years into the future versus your TotalGrowth, you will be able to compare the disk space needed without HSM to the disk space (and cost) with HSM.

At this point, you should also measure any wasteful redundancy you may have on your on-line disk. Some dormant files are simply long-forgotten files mis-placed in directories that were never cleaned out. Exact duplicates of applications or configuration files may reside in many places in your network. Advanced HSM techniques can make use of the fact that these on-line copies in different places are identical.

PHASE 2: Preparing Endusers

Before you introduce HSM, you should be sensitive to the delay HSM will cause when a user wants to access a migrated file. Although this delay may be as little as a few seconds, it will be noticed by interactive users. This

[3] "The Cost of Managing Storage on PC LAN Networks", Peterson, Michael and Neema, Farid., November 1993, Peripheral Strategies, Inc. Santa Barbara, CA.

RecallDelay is one of the central metrics of an HSM system. Also of interest is the RecallRate, the number of files that need to be demigrated in an average hour.

It's Not Just Optical Disk Any More

For practical purposes, a RecallDelay of 2 to 4 minutes or more will let you purchase affordable HSM hardware and will allow you to experiment with HSM without having to pay for the more expensive solutions needed to get RecallDelays below 15 seconds. As David Linthicum noted, "HSM systems can be priced as high as $100,000, which makes some companies believe that it may be more cost-effective simply to add disk storage."[4]

If your WorkingSet is growing nearly as quickly as your FullSet, it may not be economic to implement HSM at all. But if HSM is worthwhile and you cannot justify the cost of optical disk, it is perfectly reasonable to start your library using only tape as your nearline storage. The resulting RecallDelay will be longer, but large capacities are very economical. You can then add optical disk after you have a much more accurate forecast of your aged data usage.

Because endusers will occasionally be aware that old files will be slow, it is prudent to minimize the education they need. A responsible HSM should inform the enduser (via a pop-up window) that the file they have requested has been migrated. The user will then (depending on the features of the HSM system) be able to detach, cancel, or do nothing. If no intervention is taken, the file will be restored automatically to the location it came from.

Disk Is Neither Infinite nor Free (Yet)

During the enduser education phase, it is important to find out which dormant files are important and which files are unimportant to your users. The easiest way to do this is to prepare for your migration whether or not you already have the hardware in place.

Pre-Migration Warning Report

Some HSM systems have technology to create a pre-migration warning report automatically. It identifies all of the dormant files and prepares an email to each person telling them the list of files they own that are dormant. The endusers are thus given a chance to review the list of their files before these files are migrated. The message should also note that any dormant files the enduser needs to keep on-line should be reported to the storage administrator. This accomplishes three important functions.

(1) It heightens the awareness in the organization that disk space is a shared resource that is neither free nor infinite. Anyone whose dormant files use much more than their fair share is being a bad neighbor to the entire community. Note that migration to inexpensive media may be exactly the right treatment for those files.

(2) It gives the storage administrator specific feedback to be used in configuring the HSM systems. HSM systems with appropriate functionality can apply different migration policies to different directories or classes of files.

(3) It prepares the endusers by giving them specific examples of their own files that may become slow (if the on-line disk space is needed) after HSM migration is begun. They will be less surprised when, at a later date, they need to access a phantom file.

Note that some HSM systems lack appropriate enduser notification. For those systems, this phase is very important. Without it, endusers lacking training would be likely to misunderstand the delay imposed by recalling files. They might assume that their personal computer has crashed and re-boot.

Sizing Your Nearline Storage

Up to now, we have assumed that all of the nearline storage will be robotically available in an autoloader or a jukebox. In a multi-tier HSM, there may be an off-line tier for files so long dormant that it no longer even makes economic sense to keep them in the robotic storage. The RecallDelay for these files depends on the time it takes a human operator to mount the requested removable media. If these off-line recalls become frequent, you may want to increase the capacity of your nearline storage.

So the initial purchase of nearline storage depends on the likelihood that your endusers will need to access old files (dormant and migrated to nearline storage) and ancient files (so old they no longer fit in nearline). These two metrics depend on the frequency with which your endusers access old files. To measure that, we will compare the files that were not in the WorkingSet at time t(0) to their state now, at time t(6weeks).

[4] "Playing the Storage Shell Game.", Linthicum, David S., Open Computing, August, 1994.

If a file was dormant at time t(0) and is now in the WorkingSet at time t(6weeks), it would have been recalled if it had been migrated. The total number of these files is your estimate of the RecallRate. If this number is under 5 files per hour, you are an ideal candidate for some form of HSM.

If a file was not even in the AnnualSet and is now in the WorkingSet, the file is one that would have been among the first to migrate to nearline storage. We will use this number to estimate the number of very deep recalls that would be needed if files spilled off the end of nearline storage and had to be recalled from off-line.

If it is impractical to purchase enough nearline storage to handle your expected FullSet three years from now, make sure you purchase HSM software that can gracefully handle an off-line tier, and can properly notify users so that other subsystems do not rely on short RecallDelay. Watch out for systems with hard limits on the number of devices they can handle or the sizes of those devices.

If your current RecallRate is too high, 6 weeks is shorter than the access habits of your organization. Raise the aging parameters to 12 weeks or more until you have found a time period for which the RecallRate is comfortable. This will increase your need for on-line disk -- but will better match the usage pattern you are witnessing.

Approximate the amount of nearline storage you should start with by subtracting at time t(2years):

NearLine = FullSet(2years) - WorkingSet(2years)

Anything larger will simply delay the need to recall files from off-line storage.

PHASE 3: Adjust Aging Parameters

HSM configuration includes a high water mark (typically about 90%) that will trigger the start of migration and a low water mark (typically about 80%) where the HSM system can stop migrating data. These are percentages of your physical on-line disk.

As your FullSet slowly grows, files will be migrated to disk only as needed to free up on-line disk space. If the age (in days since last accessed) of the youngest migrated file (YoungestMigrated) drops too low, your HSM system will be overworked, recalling files that were migrated too early, running out of on-line space and then having to migrate other relatively young files.

Sometimes endusers can fill disk very quickly with new files. Advanced HSM systems will pre-stage to the nearline storage enough data to get down to the low water mark in preparation for the migration. This substantially speeds up the emergency reaction needed to prevent the on-line disk from filling up.

This pre-staging also gives you another useful metric: The age of the youngest pre-staged file. The difference between YoungestPreStaged and YoungestMigrated is the depth of prestaging. This number gives you a very good indication of the margin of safety you have between the size of your physical disk and the size of your WorkingSet of data. As a rule of thumb, if the YoungestPreStaged drops below 6 weeks, it is time to go out and buy more on-line storage.

PHASE 4: Disaster Planning

All of your storage must be accounted for in your disaster plan. The best HSM systems integrate with full, logical backup systems. With an integrated system, you would not need to worry about changes to your off-site backups.

The backup system must be aware that there is an HSM system and that a particular file is a phantom file. Otherwise the backup system would thrash the HSM system on the very first full, logical backup. For each megabyte of files that needed to be recalled, the HSM system would have to migrate a megabyte of data off of the on-line disk. This nightmare is, of course, unacceptable.

To avoid backup thrashing, non-integrated HSM systems resort to tricks like mode changing. In daytime mode, access to a phantom file causes the file to be recalled. But the HSM system can be configured to respond incorrectly during backup hours. For example, any access to a file between 2 AM and 5 AM will be assumed to be a backup access and the file will appear to be simply the tiny stub (the phantom content). Any overnight processes other than the backup will be unable to see the true data.

In addition, non-integrated HSM systems substantially complicate the rest of the backup. Either the backup system has to be told to skip the migrated data (leaving your off-site backups incomplete), or the backup system has to be able to read the data in the HSM nearline (an extremely slow process). These solutions can rapidly become impractical.

For this reason, we feel it is of critical importance that the backup system be tightly integrated into the backup and all of the other storage management functions in place at your site.

Standard Data Format

A copy of a file on removable media need not be used solely for the purpose that was intended at the time the media was written. There are substantial benefits from using the same data format for all storage management functions. For example, a tape that was originally intended to be a full, logical backup for March 3, 1995 might later be saved on the shelf because it is the best backup copy of a project that subsequently changed.

ECMA recently adopted ECMA-208 (the System Independent Data Format, SIDF) as it's suggested file and label format for removable media. This optical disk and tape format has major advantages when storing data over time, beyond just the advantage that it is the international interchange standard. ECMA-208 is suitable for storing the content, attributes and properties of files for a broad spectrum of operating systems.

According to LAN Times, "The goal of SMS and SIDF is to give both software manufacturers and end users a more stable backup platform. Although backup software and hardware manufacturers typically have touted the features of their own proprietary solutions, network administrators often found themselves stuck with one vendor for life. And if that vendor disappeared, the old media was usually not readable by other products and, thereby, useless. Enter SIDF."[5]

By storing both the backup and the HSM data in SIDF format, a common set of programs can be used to read, manage, and manipulate all of the data regardless of the reason for having written the data in the first place. In a disaster, the HSM data can be read by the backup (and restore) system. Oddly enough, it is also valuable to let the HSM system read the backup tapes. If your emergency recovery site has far less disk space than your damaged site (or if you want to recover quickly) you can simply create phantom files for all of the data that is not immediately required.

PHASE 5: Initial Manual Migration

Since HSM will introduce some delay when users access dormant files, we recommend that your very first migration be manual. This will give you a much better understanding of the kinds of files you are migrating and the ways they are likely to be used. In any sufficiently large network, it is impractical to expect a single, human storage administrator to "know" which files are important and which files are unimportant. Because the HSM system will be dutifully carrying out the storage policies, it makes sense to spend a little time formulating policies that recognize that some data is more delay-sensitive than other data.

Earlier in this paper, we described a pre-migration warning report. You may want to send out the pre-migration warning reports once a month for the first 3 or 4 months until you are comfortable with your HSM policies. Since one of the purposes of the report is to elicit feedback from the endusers, the report promises to wait one week before actually performing the migration. The process is quite simple:

At time t(PreMig), generate the pre-migration warning report. At this time the system will automatically save the list of files that were eligible. Send each user the report that shows their dormant files.

During the subsequent week, endusers will stomp, scream, whine, and intimidate you. This is how you know they are paying attention, and it also tells you which files are actually important to them. Increase (or decrease) the aging policies for those directories, wildcards, or individual files so they are less (or more) susceptible to migration than casual files. If you do not want a file migrated during this cycle, either remove it from the pre-migration list or simply access the file.

At time t(PreMig + 1 week), migrate only the files that are still on the pre-migration list and still dormant. This may be a single, built-in function in your HSM system. Although you have warned users that the files are eligible for migration, the HSM system can be told to stop at the low water mark.

Migration Policy Options

Some of the dormant files will not need to be migrated because we don't (yet) need to free up their disk space. HSM systems have policies for deciding which files to migrate first.

Some HSM systems are only capable of choosing the files to migrate based on access date. This may not be the best algorithm for you. The most configurable systems allow you to set policies based on file names, wildcards or directories that will age tools and projects differently (e.g. mark them eligible only if not accessed in 26 weeks) than casual files (which you might want to be eligible in only 6 weeks).

Two other algorithms are common. A "largest first" algorithm frees up the most space by moving the smallest number of dormant, eligible files.

The most sophisticated algorithm determines the "most eligible". This policy determines the multiple of a file's eligibility age. A file that was eligible at 6 weeks of

[5] "Backing up the SMS Way.", Arnett, Matthew F., LAN Times, July 11, 1994, McGraw Hill Inc.

dormancy is double eligible at 12 weeks. It would be migrated earlier than a tool whose multiple is 27/26 weeks.

PHASE 6: Automatic Migration

Once you are comfortable with the policies and configuration of your HSM system, you should begin automatic operations. At this point the HSM system will continually watch disk free space. As soon as the on-line disk fills above the high water mark, the HSM system will begin to free up disk space.

You may notice the nearline storage filling more quickly than you had calculated. This is because the approximations we used early in this paper don't account for the panic deletions. Endusers who run out of disk space (if you have individual quotas on each enduser) use the limit as a trigger to remind them to clean up. Since HSM is migrating their dormant files, they will only run out if their WorkingSet is larger than their quota. Since the trigger is less likely to fire, the panic and resulting cleanup are less likely.

PHASE 7: Monitoring

Choose an HSM system that fits in with the Network Management strategies you have selected. If your other systems can be monitored by Simple Network Management Protocol (SNMP), you will be able to monitor your entire network from a single console.

The HSM system will be able to tell you fullness of your on-line disks, age of the youngest migrated file, age of the youngest pre-staged file, the recall rate, and fullness of the nearline storage. If the nearline storage is full, you will also be interested in the age of the oldest nearline file.

Watch these statistics over time to do capacity planning. If the age of the youngest pre-staged file is getting too young, add more on-line disk. If the age of the oldest nearline file is too young, add more nearline storage. If your recall delays are too long, it may be time to add a new tier of faster nearline storage. Use this statistic to justify the extra cost of optical disk over tape.

LAN Magazine quotes US Connect President (Salt Lake City), Kevin Plumb, "A system that does not provide for moving files through different tiers of storage does not answer the data-glut problem; it merely moves the problem to different media."[6]

Conclusion:

An HSM system should be installed in a few, simple phases to avoid unpleasant surprises. Among those surprises are other applications that step on access dates, backup programs that are unaware of HSM, or usage patterns that make unconfigured HSM inefficient.

There are substantial differences in the HSM software available. Systems that do not provide adequate enduser notification during recall may scare endusers into rebooting their personal computers.

Systems whose policies are more configurable can give you excellent, painless control over the migration process.

HSM systems that record their data in proprietary formats restrict your ability to use that data in ways they did not anticipate.

HSM systems can complicate your backup processes if they are not integrated with backup, archiving, and your off-site protection.

An HSM system must grow gracefully and be easily managed and monitored with the rest of your network management.

Installed properly, an HSM system can substantially reduce the cost of bulk storage of dormant files with minimal impact on response times.

---

[6] "High and Mighty: Hierarchical Storage Management Increases Capacity Cheaply and Offers Scalability Without Requiring Changes to Backup Procedures.", Hinners, Bonny, LAN Magazine, June 1994, Miller Freeman, Inc.