TECHNICAL BRIEF

Application Storage Manager® (ASM) software

UNIX version ASM software customer discussion

APRIL 2004

Application Storage Manager® (ASM) software

UNIX version ASM software customer discussion

TABLE OF CONTENTS

1 STORAGETEK® APPLICATION STORAGE MANAGER® (ASM) SOFTWARE

1.1 ABSTRACT

Application Storage Manager® (ASM software) is a StorageTek® brand name that refers to a family of software products that are supported on the S/390, UNIX and Windows 2000 platforms. *The ASM software product family is defined as intelligent data movement software that is designed to automate data management and retrieval across a storage hierarchy via user-defined data policies.*

Application Storage Manager software, and StorageTek's Information Lifecyle Manangement™ strategy, allow business managers and IT professionals to blend storage technologies in a hierarchy balancing performance, capacity and costs. This is made possible by addressing two significant issues associated with data:

1) automating the data management function via user-defined data policies
2) continuous access regardless of where the data physically resides.

First, business managers and IT professionals are kept awake at night because of storage costs that are increasing in orders of magnitude with no end in sight given the current and projected data growth rates. Second, accessibility and sharing of the information asset is critical while the associated infrastructure costs (storage) are impacting profitability. Third, increasing complexity in applications, networks, servers and storage has created incredible pressure to automate the data management function to mitigate risks associated with data loss due to insufficient numbers of professionals available to monitor and administer these complex environments.

In this discussion we will examine the specific features, functionality and implementation strategies as they relate to the UNIX version of the ASM software software. While the three platforms (S/390, UNIX, Windows 2000) may share some common features, no assumptions should be drawn that a fundamental feature in any one version of ASM software will be available on a different platform.

2 DISK DATA MANAGEMENT ENVIRONMENT

Whenever there is a need for data storage in a computer environment, the discussion tends to be centered on disk and the many flavors of disk available. While a certain amount of disk (in some cases significant amounts) is needed in any computer environment, disk-only environments are not always the best choice even with the recent significant cost reductions. Business critical issues may include creating a storage area network (SAN) and a network-attached storage (NAS) infrastructure.

2.1 SAN DISK

Storage consolidation typically precedes server consolidation. A SAN disk environment is typically defined as fiber-attached disk with some significant type of disk management software. SANs are extremely important in transaction high disk throughput applications. Large database environments typically drive these needs.

2.2 NAS DISK

NAS environments are typically implemented for file serving/sharing (CIFS/NFS/HTTP), not DBMS (particularly transaction oriented). A typical NAS includes a focus on backup/recovery that includes both disk and tape. A NAS disk environment is typically defined as a central disk resource that is attached to a local network supporting a variety of network disk attach protocols. This approach may also include a SAN disk architecture.

Regardless of which disk strategy is employed, reliable, consistent and well-managed backups are essential. Disk systems can and do fail, and there are the ever-present issues with computer viruses and user mistakes (intentional or unintentional). The need for backups must be tempered by the ongoing needs of business as uptime and data processing windows continually squeeze the amount of time that is available to do backups.

Neither of these disk strategies addresses the single major issue that the majority of the data stored on spinning media is not active data and may not have been touched for many months.

Most disk systems have a constant single state (full). At a typical customer site, a data analysis revealed that as much as 80 percent of the data that is currently residing on active disk has not been accessed in the last three months. During this time, this data is taking up space, being backed up and requiring administrator support.

3 THE APPLICATION STORAGE MANAGER (ASM SOFTWARE) SOFTWARE

ASM software is intelligent data management software that is designed to automate data management and retrieval across a storage hierarchy via user-defined policies. Data management requires more than disk. The concepts of Hierarchical Storage Management (HSM) have been employed in data processing for many years in the mainframe space. HSM in the Open Systems space has not been available or sufficiently mature to implement into enterprise environments. The ASM software UNIX product from StorageTek meets the needs of enterprise data management.

3.1 CASE STUDY EXAMPLE

By analyzing the customer environment, and with the use of the "ASM software Virtual Storage Solution," customers can now balance cost, performance and capacity.

Storage, at least in the world of Open Systems, has become synonymous with disk. This is true in part because disk was easy to buy and manage, and easy to extend or add to when additional capacity was required. Talk to any systems administrator and you'll find that, as disk systems grow, the complexity associated in managing those systems also grows. Though it may be an easy decision to buy disk, it is anything but cheap.

Let's take a closer look at a business manager faced with a 20-terabyte data requirement and the choices available to solve the problem. The first thought in building the business case was to consider the traditional disk-only approach. That decision looked like this:

> 20 TB of disk	$4 million
> Backup costs	$500,000
> Mirroring as alternative to backup	$4 million
> Downtime for backups	$$$

At a minimum, this user was faced with a $4 million problem. The bigger issue was that this user faced the same decision again in as little as 12 months. In other words, there was no scalability offered in the disk-only approach.

An alternative to this "traditional" decision is made possible with Application Storage Manager. Remember, ASM software "virtualizes" storage technologies, blending performance and capacity while enabling continuous access to the information asset. A summary of the business case for the ASM software virtual storage solution would look like this:

> ASM software for UNIX (200 slots)
> ASM software Professional Services
> StorageTek D173 disk (4.5 TB)
> StorageTek library, media, drives

The total for the ASM software virtual storage solution is approximately **$600,000** to solve the same 20-terabyte data requirement. Also, given the likelihood that this business manager will need to double this storage capacity within 12 months, the incremental costs would be just over $200,000.

4 ASM SOFTWARE VIRTUAL STORAGE SOLUTION

The ASM software virtual storage solution is a combination of disk, tape, software and data management policy. The IT manager now has the ability to manage the data and where it resides rather than facing constantly growing data requirements and the growing spinning disk environment.

The Application Storage Manager software provides for:

> Enterprise architecture scalable from thousands to millions of files
> Object repository for image as well as non-image data
> Disk and tape integrated into a seamless storage solution
> Applications insulated from the storage and associated management policies
> Data backup and disaster recovery costs reduced

The network attached storage (NAS) concepts that were introduced into data centers with disk only have been expanded into a total storage environment. No longer does a system administrator have to make the decision of how long to keep the data and what must be deleted from spinning disk. The ASM software environment combines the best of disk and tape into a seamless storage environment.

Data is stored into the ASM software environment via a number of available network storage protocols. This data is stored with a known file name that is used as the access method for the data. Internal knowledge of where the data is located, etc., is virtualized to allow for the best method of storage based on a number of storage policies that are defined and maintained by site system administration personnel.

5 WHERE IS ASM SOFTWARE UNIX INSTALLED TODAY

The Application Storage Manager UNIX software product is currently installed at over 600 sites worldwide; it is currently installed in the following disciplines:

- Financial services
- Check image archives
- Medical Imaging (radiology (PACS), cardiology, enterprise data) archives
- Energy — oil and gas exploration
- Scientific (research and trials)
- "Rich Media" — security surveillance

6 HOW DOES ASM SOFTWARE MAKE MY JOB AS SYSTEM ADMINISTRATOR EASIER?

The ASM software software automates the task of managing data in the storage environment by implementing storage policy around the major functions of Application Storage Manager software. The storage policy may include archiving, releasing, staging and recycling.

6.1 ARCHIVING

Data is written to the ASM software storage environment. Once the data file reaches a predetermined ideal age, the ASM software system enforces the defined data management policy. Options include:

> The ability to generate up to four copies simultaneously or on a delayed basis. Each copy can be on unique media to take advantage of the best data storage environment.
- Very reliable, very fast tape
- Slower data-dense tape
- Remote tape libraries
- SSP specifically set up to remotely store ASM software-generated data
- Remote disk farms
- Data isolation
  - – Data is stored in the most cost-effective form.
  - – Tape (current method; future may be some other technology)

> Data written onto media in "tar" format. This allows data to be accessed from the storage media without ASM software in the event of a disaster

> Archive sets and families can be freely and easily defined

> Optimized archiving of large and small files
- Fast seek times
- Fast transfer times

6.2 RELEASE

> Release only if a valid copy exists

> Release directly after archiving is an option that is often used when enterprise backup data is stored into an ASM software archive

> Release at watermarks

> Never release data

> Data stays on fast-access spinning disk until the storage resource is exhausted

6.3 STAGE

> Associative staging per directory or selection of files

> Files with associative attribute set will be staged together

> Read-behind stage, data immediately available to users or applications while staging of files in process

> Never stage — don't write data onto disk, data will be given to user directly

6.4 RECYCLE

> Search for media that contains data that is no longer needed. Active files are marked for re-archiving
> Recycler marks active files on media for re-archive
> Active files are re-archived to new media
> Archive marks old media as no longer valid
> Recycler reclaims old media after no more valid files are left on media

7 CONTINUOUS BACKUPS

As files reach their idle timer, a copy of the file is put on tape as part of the data policy. As the file is subsequently changed, all of the deltas are also archived to tape. This allows for the ability to retrieve multiple versions of a single file.

For data backup needs that do not lend themselves to Application Storage Manager system control of all the data, there is the method of using an enterprise backup system (Legato, VERITAS) in order to have the backup data written to an ASM software disk partition that instantly moves the data to tape and then releases the data from the disk. This allows for a single enterprise data management policy as it relates to disk and tape without having to completely redesign current practices in the data center.

8 ASM SOFTWARE DISASTER PROTECTION

> File logging
> Archived files are written in "tar" format
> ANSI standard volume labels
> Directory and file location info is also archived
> Recover even largest file systems in minutes
> Files can be recovered with or without ASM software

9 TECHNOLOGY OBSOLESCENCE PROTECTION

As data environments grow, the value of the data begins to outlive the storage medium that it is stored on. Traditional methods of migrating the data to a newer storage environment often involved tremendous loads on the applications that generated the data. This is true anytime that there is a system that records the physical location of the data on specific media and its location on that media whether optical, tape or disk. The question is no longer, "is the media good for 100 years?" but, "how long will the data be able to be read on that media?" There are a large number of data storage archives that have 14-inch, 12-inch, 5-inch MOD, 5-inch WORM and now DVD optical archives. The same can be said for tape archives as tape is advancing (9-, 18-, 36-track, DLT, SDLT, LTO) every bit as quickly as computers, memory, printers, CPUs and etc.

> ASM software provides a layer of separation from the data and the storage technology
  • This allows for the use of the most economical storage solution with a single interface
> In the past this has lead to a performance hit
  • With properly sized disk cache this is no longer an issue

The question now is, "when I sell you a long-term archive, what is the plan/design to move to the next generation storage technology?" and, "what are my options for divorcing a storage solution (optical, tape, SSP, deep archive, etc.)?"

By placing the data in a virtual storage environment where the application records the file name as the access method, with no knowledge of the actual storage location, a large world of options is available for the long-term management of the data. This separation of the data management policy from the data generation allows employment of data storage best practice without regard for what specifics the data generation application supports. Data is moved as a background function. No load is placed on the data generation server.

There have been issues with the data generation server not being able to keep up with the technology churn (i.e., application does not have the latest drivers or support the current generation of storage devices). As technology moves forward, it is often a trade-off. "Do we put additional resources into the data generation and analysis portion of the application or put additional R&D in the archive that has a minimal value in the sales process?"

> Transparent migration from old to new storage technology
  - The storage technology of today will change

> Rapid certification of new technology

> Data archive growth as needed

> Customer investment protection

> Provide a proven method for migration to next generation technology

10 MULTIPLE-USE ARCHIVE

The tape library can support many data streams simultaneously. An archive originally designed to serve as a cardiology archive can be expanded to support radiology and IT enterprise backup. This allows less total hardware to do more.

> Provide a method to allow multiple data streams into a single archive
  - Strategic advantage to allow cardiology and radiology in a single library
  - Site already has an ASM software environment; add your solution at greatly reduced cost

Dedicated standalone digital archives have been the standard level of practice for many years in the health care enterprise. This is due more to historical reasons than any other. PACS providers would provide only a single archiving option and the departments had a 'take it or leave it' option. These archive options were typically a SCSI-attached archive that was connected directly to the PACS database server. Data stored in the archives was typically in a closed proprietary format and could be accessed only through the PACS access interface. This is still very much the case with the standalone tape drives that are provided with each MR and CT scanner today.

Information islands can also be seen in several departments throughout the health care enterprise as they maintain specialized databases on their patients. These include neo-natal growth charts and risk classifications, pharmacy databases, pathology databases and labs. Each of these servers represents a data island and

a potential point of private patient data that could be mistakenly released to the public. They also represent a critical health care enterprise activity that needs to have a data policy in place for the care, disaster protection, access control and patient protection procedures.

Once the long-term archive is freed from the direct control of the PACS archive, it becomes a sharable resource through the use of ASM software (or at least through a library sharing software such as StorageTek's ACSLS Manager™ software, LibraryStation™ software) and a number of additional opportunities and options become available. The health care enterprise can begin to realize the value of having a true enterprise sharable archive. Some of these benefits include:

> **Greater cost-effective storage.** The age-old adage is very apparent here: the larger the package the cheaper it is. The cost per megabyte storage is much less the larger the archive library.

> **Reduced headcount.** A single part-time administrator can monitor and maintain a central storage resource much easier than having archives distributed throughout the health care enterprise.

> **Each department gets more than they paid for**. For example, if three tape drives were needed to perform the enterprise backup in a given time window, radiology needed three drives to perform the pre-fetch and peak daily ad-hoc retrievals and cardiology needed three tape drives for their peak daily needs. A tape library with a total of six drives would be able to perform all these activities and give greater throughput and redundancy. Again, if a tape library had three drives and one was unavailable because of maintenance, one-third of the capacity would be lost and the backups would take 33 percent longer. In a larger shared archive, the loss of a single tape drive represents a loss of one-sixth of the total capacity and still makes a sufficient number of drives available to complete the required tasks on time.

- 18:00 — 21:00 Cardiology batch store: three drives needed
- 22:00 — 1:00 Radiology prefetch window: three drives needed
- 1:00 — 4:00 Enterprise backup: three drives needed

In a traditional captive archive there would be three libraries and a total of nine tape drives. In a shared archive with a total of six drives, three tape drives would be dedicated to the time-based task with three available for down drives and emergent requests for archived data from the other services.

> **Greater security.** A single shared enterprise archive can have additional security procedures and access controls put into place that would otherwise be cost prohibitive if they were distributed throughout the health care enterprise. If libraries are distributed (as they traditionally have been), there are many more possible points of security failure.

> **Data isolation.** By using a single archive strategy for the health care enterprise, a policy based storage strategy can be employed and enforced.

> **Maintain data access viability.** This allows the migration of the archived data to the next generation storage technology in a controlled, proven process.

Once a shared health care enterprise archive is implemented, what additional benefits can be gained?

- A true inter-department resource where no one department has to own the archive. Economics of scale are possible that could not be achieved before.
- Standard IT practices can be implemented in the health care environment. Disaster recovery planning is possible and practical.

11 TOTAL COST OF OWNERSHIP IN A LONG-TERM STORAGE ARCHIVE

The total cost of any data archive system includes much more than the initial purchase price. Some of the components of ownership include:

- **Initial equipment purchase.** This is the most obvious and is often given the greatest amount of scrutiny when an archive is considered. It is also one of the easiest targets to attack to reduce the perceived cost of the archive. We have seen, however, that short-term reductions in the initial cost can lead to very large expenses later in the life of the archive. (Migration costs are a major event that must be planned for from the beginning.)
- **Maintenance.** This is the cost to maintain an archive with respect to monthly hardware maintenance, downtime for preventive maintenance and unexpected failures.
- **Media**. What is the piece cost per megabyte? As future growth plans are reviewed, can the media be reused or must it be replaced?
- **Servers.** What is the cost to maintain, monitor and connect to the archive servers? Is a single server more cost effective than several? Does a single server serve the long-term needs of the health care enterprise?
- **Data protection.** What is the ultimate capacity of the archive? Will several be needed to meet the anticipated growth needs? Will media have to be put on the shelf with an operator taking the role of the archive?
- **Data obsolesces (migration).** This is often the last thing considered when an archive is installed. What is the exit strategy? Is there a plan for moving from the current technology to the next generation? What is involved and has it been tested?

12 WAYS TO CONNECT TO ASM SOFTWARE SERVER

The Application Storage Manager software provides HSM disk management to a UNIX file system. HSM allows for a disk to appear to the user as a resource that is much larger than actual physical disk. As the disk fills to a predetermined high-water mark, data is removed from the disk, while directory information is maintained. As the data that has been removed from the disk is requested, the ASM software software then causes the data to be restored from tape back to the disk. The UNIX file system is then available to receive data and migrate it to/from tape.

This UNIX file system can have data placed on it in a number of ways; each way has advantages and disadvantages.

12.1 NAS

> Integrates with standard NTFS partitions without changing any NTFS attributes
> ANSI Standard Volume Labels
> Backup software aware so directory and file location information can be backed up with standard NT backup packages
> Recover within minutes using standard backup package restore options

12.2 NFS

> This is generally the quickest way to move data from a UNIX system to the archive.
> NFS is the native method for a Sun Server to distribute disk resources.
> When transferring files to and from ASM software via NFS mounts, remember: NFS data is written in 32K blocks. This means that a 100k file is four blocks with the final block containing 28k empty data. Each block is receipt-notified to the sending machine, which is an additional 32K block. So for a single 100k file an extra 50k of overhead is incurred on the network.

12.2.1 Advantages

The HSM-managed disks appear to the world as a single very large disk. Generally very little modification is required in the archive application to take advantage of the connection.

12.2.1 Disadvantages

> NFS mounts do not work well over WANs (wide area networks).
> Large amount of overhead for each packet of information
> NFS is prone to lockups on large data transfers with Solaris 2.5.1

12.3 FTP

FTP is becoming a very popular method of transferring files between dissimilar architectures. The server machine packages the files to be transferred into a single large file and then sends them to the tape repository. When the data is needed to service a query, the server requests the data package back, dissects it back into the native components and serves the needs of the user.

12.3.1 Advantages

Works very well over WANs

12.3.2 Disadvantages

Very little error detection/correction; possible to get zero length or incomplete files. Verification of complete/correct file transfer is a must.

12.4 RCP

RCP works very well in a UNIX to UNIX transfer.

12.4.1 Advantages

Works very well over WANs

12.4.2 Disadvantages

Tends to be a difficult protocol to implement on PCs.

12.5 NT FS

Samba from Sun and other vendors in addition to other connectivity software (Solaris PC NetLink, or Solaris Easy Access Server) that allows the local disk on the Sun Server to appear to the NT world as an NT file system. Solaris PC NetLink is a software product that allows Sun servers to run native Windows NT 4.0 network services on the Solaris operating environment software.

Solaris PC NetLink is a key addition to Sun's Windows NT interoperability product line. Based on AT&T's Advanced Server for UNIX products, Solaris PC NetLink provides transparent connectivity into Windows NT network environments, allowing customers to replace Windows NT servers with the more reliable and scalable Sun solution.

12.5.1 Advantages

UNIX file system looks like an NT server to NT/95/98 clients

12.5.2 Disadvantages

Version 1.0 tended to be slow; this appears to have been corrected in the current versions.

13 GREAT IDEA, BUT WHO SUPPORTS THE ASM SOFTWARE ARCHIVE MANAGER IN THE HEALTH CARE ENVIRONMENT?

The ASM software software has been implemented or is currently supported by the following PACS providers:

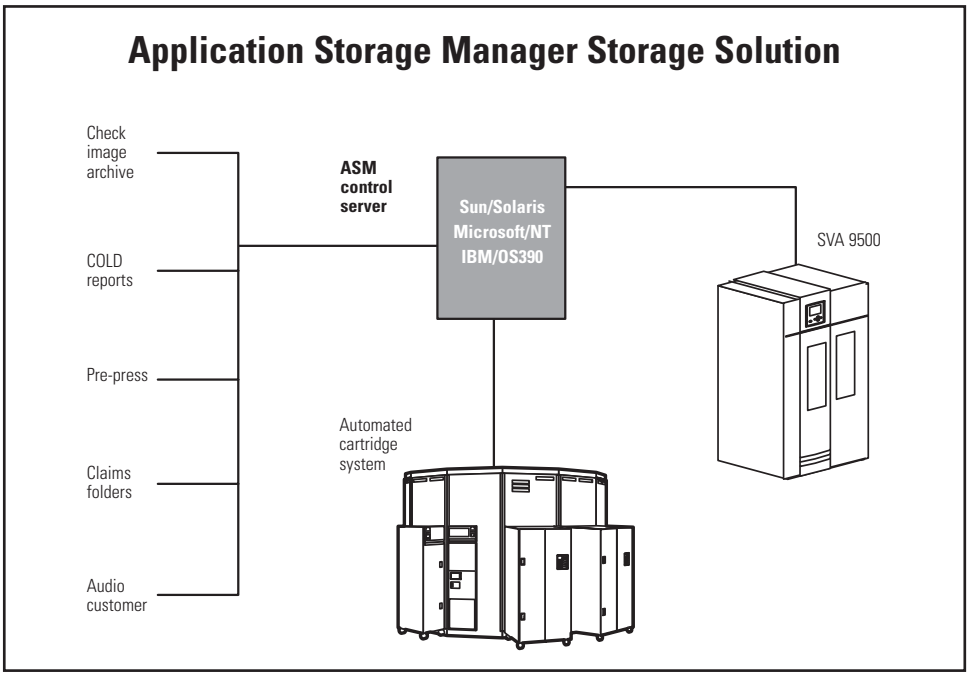
> AGFA: Agfa Healthcare
> A.L.I. Technologies, Inc.
> Emageon
> Philips Medical Systems
> Siemens Medical Solutions
> Witt Biomedical

ASM software allows the selection of the best-in-class technology for the application; today this includes disk for short-term archival of image data (in some cases this may be one to two years). Then a combination of longer-term storage and backup strategies that include access centric tape for longer-term storage of images with little degradation in retrieval time and data dense tape for disaster recovery/business continuity.

14 COMPLEX APPLICATION ENVIRONMENTS

The illustration below is a simple example of the many complexities facing system administrators today. Applications may be hosted on different types of servers, network speeds and topologies may vary, and the direct-attached storage approach is so expensive that it is no longer a viable and sustainable practice in many organizations. The StorageTek Application Storage Manager virtual storage solution is a network-attached storage system that integrates a multi-tiered storage hierarchy allowing the customer's data to be in the RIGHT PLACE at the RIGHT TIME with the RIGHT PERFORMANCE for the RIGHT PRICE.

Because ASM software software automates the data management function, many different applications can share common "pools of storage." Data policies that control how images may reside on disk, or when those images are retrieved from performance tape or archive tape, may be quite different from the data policies managing the "claims folders" application data. This approach actually introduces a level of simplification and automation that otherwise would not exist. In this sense, Application Storage Manager software and the virtualization of storage technologies become an integral part of the customer's business operations. In addition, this solution is designed to increase the availability of the information asset to the customer's various business units while having a very effective price/performance ratio. It offers the potential of eliminating the need for microfilm/optical as a storage medium, provides support for multiple archive sites and multiple concurrent archive users, and is targeted as one of many of the business continuity issues facing our customers today.



15 SUMMARY

Along with its best-in-class partners, StorageTek has the ability to offer complete storage solutions incorporating disk, automated tape and data management software in the way of ASM software. These solutions can be offered as individual components and/or can be offered in direct attached, network attached or SAN-enabled configurations. Perhaps best of all, these storage solutions are scalable from gigabytes to petabytes. The end result is that no one can help their customers control cost and address such a wide range of data and storage issues better than StorageTek.

STORAGETEK HEALTHCARE PACS PARTNERS:

Agfa Healthcare

Philips Medical Systems

Rorke Data, Inc.

Kodak Healthcare Imaging

Siemens Medical Solutions

ABOUT STORAGETEK®

Storage Technology Corporation (NYSE: STK), a $2 billion worldwide company with headquarters in Louisville, CO, has been delivering a broad range of storage management solutions designed for IT professionals for over 30 years. StorageTek offers solutions that are easy to manage, integrate well with existing infrastructures and allow universal access to data across servers, media types and storage networks. StorageTek's practical and safe storage solutions for tape automation, disk storage systems and storage integration, coupled with a global services network, provide IT professionals with confidence and know-how to manage their entire storage management ecosystem today and in the future.

StorageTek products are available through a worldwide network. For more information, visit www.storagetek.com, or call 1.800.275.4785 or 01.303.673.2800.

WORLD HEADQUARTERS

Storage Technology Corporation
One StorageTek Drive
Louisville, Colorado 80028 USA
1.800.877.9220 or 01.303.673.5151

MS 9144 D 04/04