
Introduction

This document provides an overview of the storage management product *Archimedia* from industry specialists El Fresko Technologies. The report describes the key functionality and benefits of the product and discusses the technology in greater detail.

It also clarifies some of the issues raised that potential customers should consider when buying archiving systems and lays down some guidelines for the most appropriate product to be used in any given situation.

Archimedia is supported on Windows 2000/XP/2003 and Sun Solaris.

Background

System managers and computer users are increasingly being forced to re-examine their long term archiving and data storage policies in the light of constantly falling hardware prices and new operational requirements. One overriding concern is the ever increasing amount of data which must be made immediately available *on-line* (i.e. within a matter of seconds). There is also the requirement to store or archive large amounts of data for long periods of time, and the related need to store data that cannot be amended for legal reasons, such as accounts and insurance information. Together these factors are forcing users to look at technologies, such as optical, CD recordable (CD-R) or Digital Versatile Disc (DVD-ROM/RAM/R) to solve their storage and archival needs.

Many users have already realized the benefits of CD, and optical disk technology, for cost effective long-term storage of valuable computer data. They have found a way to remove the need for operator intervention and ensure reasonable user response times using jukebox technology. Library systems (often called jukeboxes) are ideally suited to meeting these requirements because of the robust nature of the robotics, the massive amount of data that can be stored and the guaranteed lifetime of media (typically over twenty years).

DVD-RAM/R technology offers much larger capacities than CD-R which means that the cost per Mbyte of storing data is now much lower.

A jukebox or library system can be used to store anything between many tens of Gbytes of data up to several Terabytes. Access times are typically very fast and a file can be retrieved in seconds rather than the hours it can take to load and restore data from an off-line tape.

Hardware is not sufficient on its own to provide complete archiving solutions. Without the availability of sophisticated and reliable software to ensure it is efficiently utilized, the hardware is very limited.

El Fresko Technologies are specialists in the area of mass storage system integration and supply a number of products designed to provide network-wide archiving and data management solutions. This document discusses the features and benefits of *Archimedia*.

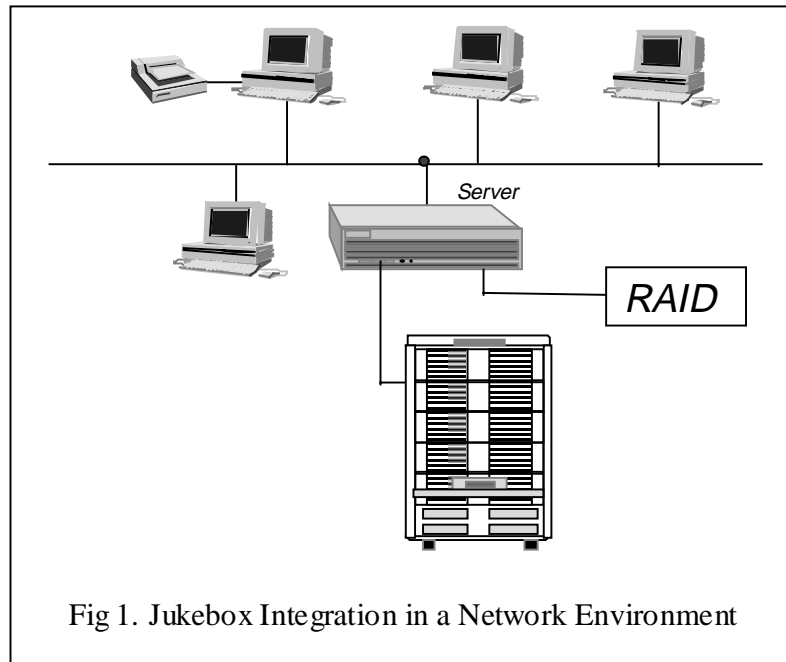
Current Jukebox Management Software

Over the years many users have discovered that it is not enough to simply purchase a jukebox and attach it to a server/network to have a working and manageable archive. To get reasonable performance and the ability to use systems in 'live' environments where tens or even hundreds of users are accessing the archive they require something else: sophisticated jukebox and storage management software.

Much of the jukebox management software on the market today can be classified as *toolkit* software, *disk emulation* software or *Hierarchical Storage Management (HSM)* software. All of these approaches have their benefits but rarely do they provide all the answers for users wanting to store enormous amounts of data on-line for long periods without the need for constant supervision.

Toolkit software is designed for software programmers so that jukeboxes can easily be integrated into application software. It has little if any use for system managers. It usually consists of a set of libraries and utilities that must be called from within other applications or from shell scripts. Toolkit software often lacks any robotics error recovery capability or key management tools like performance monitoring and graphical user interfaces.

Disk emulation software has proved very popular over the years. It allows each disk in a jukebox (or more exactly, each side of each disk) to be viewed as a file system device. By integrating a jukebox with the operating system via the file system it is possible to use third party applications directly, without the need for further software.



However, disk emulation software has a number of drawbacks especially in a loaded networked environment. In such environments it is possible to have a large number of read/write requests that cannot be serviced immediately. If there are two drives in the jukebox and there are three requests for data then one request must wait. This may lead to the jukebox starting to move discs in and out of drives in a rapid and seemingly uncontrolled manner (known as thrashing). In a network, user file requests are repeated at regular intervals and the whole network can become swamped with such requests (which results in the frustrating but well known message: “*Server not responding*”).

HSM systems offer to solve all these problems but users are finding these solutions over elaborate for what they are trying to achieve. Also the systems are so complex that when something does go wrong users are not able to quickly and easily diagnose the problems themselves. We discuss HSM systems later in this paper.

El Fresko Technologies designed *Archimedia* specifically to overcome these problems. It combines the latest caching algorithms with a powerful volume management system, configured by an intuitive graphical utility to provide an integrated jukebox solution in a network system.

Archimedia - Sophisticated Archive Data Management

How Does Archimedia Operate?

As already discussed, for many users simple jukebox control software on its own is not sufficient. Users require additional features such as fast network performance (which implies file caching), support for media that is held off-line, data mirroring and built-in fault tolerance. These features are all provided as standard in *Archimedia*.

Archimedia overcomes many of the problems normally associated with jukebox systems, such as slow response times, primarily by caching key information on the hard disk.

By integrating hard disk or RAID technology in this way a jukebox becomes the ideal tool for the data archiving environment where massive amounts of data must be held on-line at any one time.

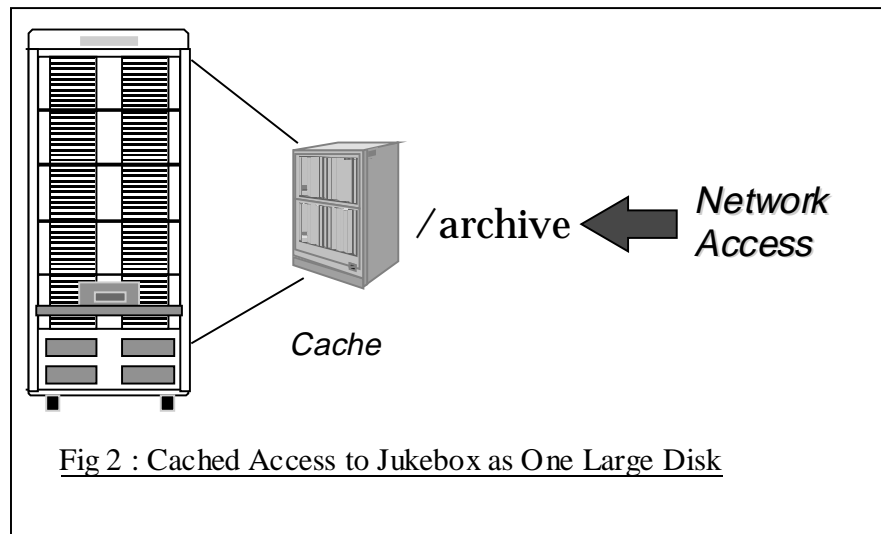


Fig 2 : Cached Access to Jukebox as One Large Disk

Secondly *Archimedia* allows a jukebox to be viewed as one large file system, making the management far simpler. It has been developed over several years and is being used in some of the largest optical jukebox sites around the world.

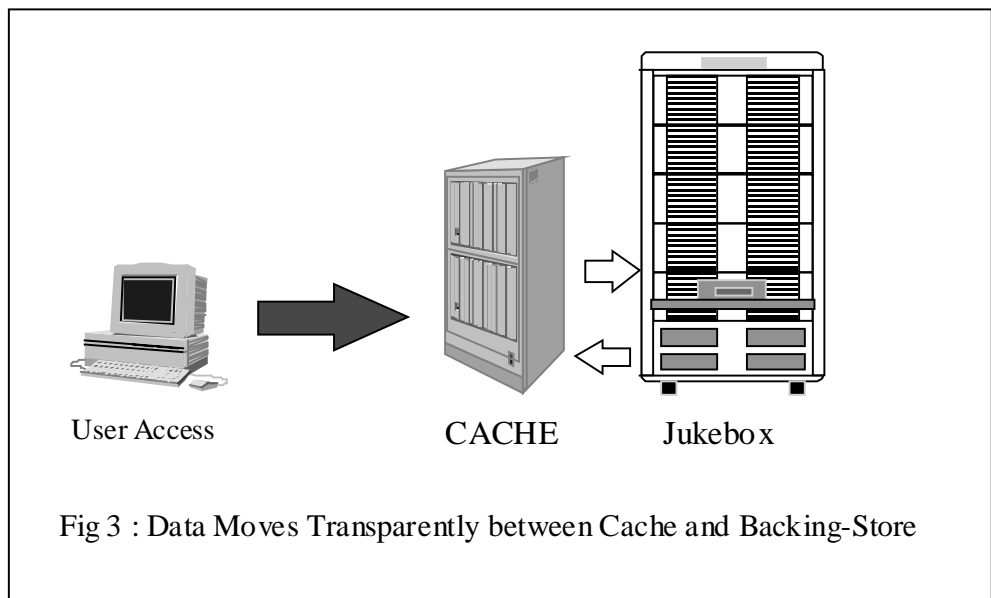
Archimedia is an integrated solution combining several logical modules: volume management, data and directory caching components, the established Jukebox Control Software, together with a user friendly graphical interface. It also includes support for WORM and rewriteable MO/MF, as well as DVD-ROM, DVD-RAM, DVD-R, CD-ROM (read only) and CD-R (CD recordable) library systems.

Data Caching

There are enormous advantages in ‘caching’ key information when using DVD, CD or optical disk technology, since these devices are much slower than hard disks. El Fresko have developed a general-purpose directory and file caching system for use with optical, DVD, and CD, which has been improved over a number of years, to provide a configurable high performance system.

It can handle millions of small files in hundreds of separate directories as easily as thousands of large files in a simple flat directory structure. It has been optimized for a wide variety of environments such as CAD/CAM, data logging, document management, Pre-Press printing, as well as data archiving.

All the caching software is provided as an integral part of the *Archimedia* product. Different aspects of the caching software are explained more fully in the sections below.



When the file is accessed (either by reading from it or writing to it) it is transparently and automatically copied into the file cache on hard disk. Subsequent reads can be serviced immediately from the hard disk, freeing the jukebox drives for other operations. This can considerably improve the speed of read requests and reduce the load on jukebox robotics and drives.

If a file is being written it will first be created and built up in a cache. After the file has been idle for a predefined length of time or in response to a user request it

will be written or 'flushed' to the file system on the disk. By batching up write requests in this way it is possible to maximize the burst of speed of the drives.

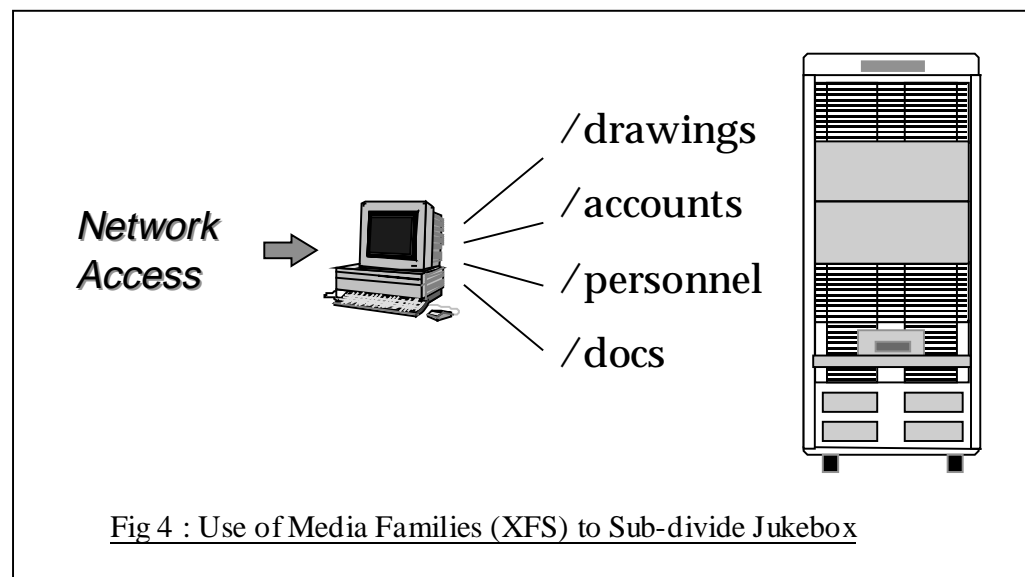
The size of the cache for each *Archimedia* file system is configurable at system set-up time and can be altered to meet requirements of individual sites. The size can vary from anything from a hundred Mbytes to tens of Gbytes and it can be dynamically increased without having to reboot the server. The cache can also span multiple discs if required.

Logical Volume Management (or eXtended File Systems - XFS)

The *Archimedia* software allows disks to be combined together as a *media family*, called an XFS, which appear as a large, single file system. In a Windows system they appear as a drive letter; in UNIX as a mount point. An XFS is mounted in the normal manner (via SMB or NFS) and data can be streamed to it using standard utilities. *Archimedia* limits the number of mounted file systems that are in use which considerably reduces the amount of system administration required.

It should be noted that the file system on each disc is in standard format, NTFS, UFS, UDF or ISO 9660 depending on the host Operating System. Each disk is therefore readable using a standalone drive, without the need for any third party software.

For instance a jukebox can be partitioned to hold an archive of images that are rarely accessed but which must be held on-line so that designers can always browse old images.



In the diagram above we show a site which has divided the jukebox into four separate file systems. Because CAD/CAM drawings tend to be large, half the media in the jukebox was allocated to this file system. As each file system is entirely separate they can be shared with different permissions, for example Engineers can be given access to 'drawings' and 'docs', but restricted from accessing 'personnel'; the Finance Department can be given full write access to 'accounts', read access to 'personnel' and no access at all the to the engineering data.

Each file system can also be mapped to different drive letters, for example **p:** for the personnel data, and **q:** for the accounts. *Archimedia* is completely flexible in the manner in which it allows the jukebox to be partitioned and accessed.

There are a number of benefits to be gained from using the extended file system capability of *Archimedia*. The primary reason is that it takes away much of the manual intervention required by a system manager when setting up and maintaining a system. The user and application developer need no longer be concerned on which disc a particular file resides. The software ensures that when one disc fills up the next in the sequence is used. The algorithm used to select which disc in a volume set to use is defined when the system is configured.

Each disc in an XFS is an integral file system in its own right and can be accessed in a standalone drive without any extra software. Also this system reduces the load on the host operating system by only mounting a file system when it is actually needed rather than mounting each disk at boot time.

In addition to the mounting of an XFS as a file system, it is possible to access a volume set directly using special utilities provided by El Fresko Technologies. This is appropriate in some situations, such as when the archive is first set up, since the utilities are much faster than going through the standard network interface.

El Fresko Technologies' aim is to provide a general purpose but flexible package designed to meet the different needs of various classes of users: system managers, systems integrators, application developers as well as general end-users.

Off-Line Media Management

Once the jukebox becomes nearly full of data it is necessary to consider the options. If the data is no longer relevant or has become redundant, it may be sufficient simply to reformat the discs in the case of magneto-optical, and then reuse the same discs again. Alternatively it may be necessary to purchase another jukebox or to investigate the possibility of taking some discs out of the jukebox and storing them off-line, usually in a protected area such as a fire-safe. El Fresko Technologies supports all these options and we have spent considerable time talking to users about issues concerned with off-line storage.

Archimedia is able to manage the export of selected discs or a whole XFS from a jukebox. The system holds in the directory cache details about all files which were once in the jukebox. When a user tries to access a file, which is now off-line, the system will generate a message informing the user that the data is not available. It will also inform the system manager that someone has attempted to access a file on an off-line disc. The administrator then has the option of loading the disc back on-line (either in the jukebox or a stand-alone device) or taking alternative action.

Backup and Restore

Since all data controlled via *Archimedia* is stored in a standard file system format, existing utilities can be used to back up portions of the jukebox to tape and other media. Also the directory and the file cache information are in standard UNIX/NT format and can be backed up in the normal way.

The back up of every disc in an optical, DVD or CD jukebox can however take a considerable length of time, even for a simple incremental back up since each disc (or large number of discs) will normally need to be loaded into a drive at least once during the back up process. In fact existing back up programs and scripts may start to thrash the jukebox since they do not know which discs contain which files.

Archimedia has extra functionality built into it to overcome this problem. It knows which files reside on each disc so that a whole disc's worth of data can be backed up in one pass. It also keeps track of all data that has changed *through the cache* over a given time period and then only needs to back up these files. El Fresko Technologies provides a number of special utilities (based on standard BSD software) to intelligently back up both the data in the jukebox itself and also the control data.

Note that when using *Archimedia* (unlike some packages) the control data can be easily reproduced from the discs themselves. This may take a considerable length of time, particularly for large XFS file systems and we recommend that the control or meta data is backed up regularly as well. Once the control data has been restored then the archive can very quickly become useable again.

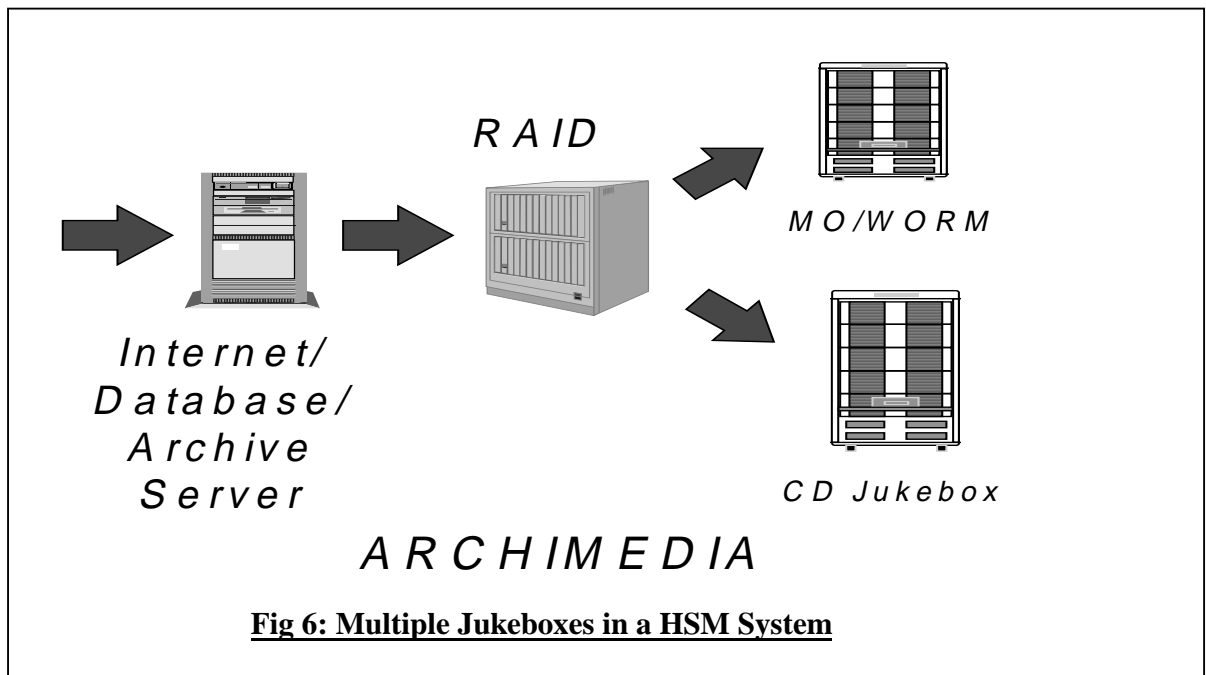
Hierarchical Storage Management

When people talk about a Hierarchical Storage Management (HSM) System they normally mean a system that automatically 'migrates' or moves data from one media type to another, more appropriate media type based on well-defined criteria (such as date/time or last used time). This takes place either at a set time or when a hard disk starts to run out of space. In most cases this means the HSM automatically moves data from a hard disk onto optical, and finally onto tape. The HSM software ensures that the data is moved 'forward' and restored to the local hard disk when a user or application accesses the data.

HSMs keep an on-line directory of where all the data resides and many have built-in fault tolerance (such as mirroring) in case any part of the system fails. An HSM initiates the process of automatically moving data from hard disk or optical onto off-line storage (e.g. tape or CD) and the 'back-up' of the HSM itself.

HSM has also come to mean the process of automatically moving data from the local client's hard disk to the network server's hard disk as the local disk fills up. These are sometimes called Network HSMs (NHSM) or distributed HSMs (DHSM)

Note that this description covers two distinct but related activities: one is the archiving (or moving) of data to a different media and the second is the freeing up of the hard disk space. No wonder users are confused when discussing types of hierarchical storage management!



The *Archimedia* family contains many of the features of HSMs, such as: transparent control of a jukebox; automatic movement of data from CD, DVD or Optical disc to hard disk; on-line directory information; and support for disc spanning. In fact *Archimedia* may be far more appropriate and provide a more cost effective solution since the full functionality of HSM may not be required.

This is particularly true where a system is only being used to archive data, rather than to free space on a number of servers and clients. El Fresko Technologies' approach has been to give the System Manager and the computer user the ability to control their own data migration policies using standard UNIX tools and established applications. Thus a user can copy data from a hard disk to the *projects* directory (F:) on the archive server and know that this information will be available tomorrow or in several years time. If the data is required it can be examined on the hard disk then explicitly copied to a local disk. This does not happen automatically but there is no need to wait for an operator to load a disc to restore a file.

The provision of standard data management tools adds more flexibility and is simpler to understand than sophisticated hierarchical data management products. This means the system manager need not learn extra commands but can quickly configure and optimize the system to meet specific site requirements.

Features of Optical CD and DVD Storage Media

This article does not discuss in any great depth the features and benefits of optical or CD-ROM disc technology. Information is readily available from the manufacturers and suppliers of the hardware, such as JVC, Hewlett Packard, Pioneer, Plasmon, Maxoptix, NSM, Sony, to name but a few.

However CD/CD-R, DVD-ROM/RAM/R and MO/WORM media offer enormous advantages over other secondary storage devices such as tape, microfiche and paper. The CDR and MO media are now guaranteed to well over twenty five years, with manufacturers such as HP now guaranteeing MO/WORM media lifetimes for fifty years, based on accelerated testing techniques.

Media lifetimes of CD-R discs have been the subject of much discussion recently with some sources claiming that discs may become unreadable in five years. However the manufacturers are keen to deny this and Kodak will offer a one hundred year guarantee on their Infoguard CD-R discs.

Features of CD-ROM and CD Recordable

International standards such as ISO 9660 and Rock Ridge meant that CD-ROM discs could be read on virtually any computer (e.g. PC, Apple, UNIX). Nowadays factory produced silver CD-ROM is the *de facto* standard for the distribution of software and a wide variety of computer-generated data such as catalogues and patents.

The take up of CD recorder technology has also been impressive and CD-R drives already out sell MO and WORM drives. As the price of the drives and the media is set to fall even further this trend is likely to continue.

CD-R drives allow users to create *multi-session* discs (CD-XA) which can be very useful for archiving, since it is possible to build up a set of time stamped 'images' of the system over a period of time. Unfortunately CD-XA discs are not always readable by default on all systems without the availability of special software.

One main limitation this technology is that CDs must be recorded in 'real time' without interruption to the flow of the data during recording, otherwise the discs become corrupted. Availability of high performance components and software such as El Fresko Technologies' *CDcreation* product has reduced this risk but users are still sometimes frustrated by the need to make CD's in batch mode.

CD technology has achieved a very high level of acceptance in computing, particularly in sectors such as imaging and printing, not only because of the low cost per Mbytes but also the widespread adoption of standards.

Features of MO and WORM

MO and WORM technology is also well established in a number of particular market sectors where they offer significant advantages over CD-ROM.

The key features of optical disc technology are:

- Long guaranteed media lifetimes
- The very high capacities of optical discs and library systems and the scalability of the units. Entry level systems start at 5.2 GB (now 9.1 GB drives are viable) but are virtually limitless when large jukeboxes are combined
- Data can be retrieved very quickly from optical discs. Seek times are very fast and access times are comparable with hard disk performance
- The reliability and robustness of the optical disk drives and jukeboxes is very high. Mean times between failure (MTBF) usually exceed 80,000 hours and with a mean time to repair of less than 30 minutes, this shows the technology is now very reliable compared to other autochangers and stackers

Write Once Read Many (WORM) technology is particularly useful when the user must ensure for legal or operational reasons that the data can never be amended or deleted. WORM technology has been available for over ten years now. It is used in a wide variety of environments such as banking, financial and legal areas where it has become the preferred solution.

Features of DVD

Digital Versatile Disc technology is the latest development from the manufacturers promising much larger capacities than CDROM and at much lower costs.

The key features of DVD technology include:

- Increased capacity from 2.6 GB per side rising to over 19 Gbyte per disc
- Much faster performance (>1 Mbyte per sec)
- Support for Write Once (DVD-R) as well as re-write (DVD-RAM / DVD-RW)
- Widespread acceptance of hardware and software standards

DVD overcomes many of the shortcomings of CD and utilizes many of the advanced features of optical discs.

Most standards have now been accepted for DVD technology including those for factory produced DVD-ROM, write-once DVD-R and read/write DVD-RAM. DVD-RAM drives offer true re-writability of discs on a per sector basis.

The file system format for DVD is defined by the ECMA standard and is called the Universal Disk Format (UDF).

As costs of DVD drives and media continue to fall, driven by the mass consumer market and DVD is expected to replace CD-ROM/CDR and optical disc technology in the next one to two years.

Choosing the Correct Media Type

Each type of media has its own merits and is suitable for different requirements. In some cases you will be limited by the hardware which may already have been selected.

There are a number of issues to consider when selecting the most suitable media type.

- If you simply want to access pre-written CD-ROMs or DVD-ROMS then you should select a CDROM file system (read only) and CD or DVD drives.
- If you want to archive data then the choice is between CD recordable (CD-R), magneto-optical (MO), Write Once Read Many (WORM), DVD recordable (DVD-R) and DVD-RAM.
- In terms of cost, the cheapest solution per Mbyte is still CD, followed by DVD and then 5.25" MO with 12" WORM being the most expensive.
- In terms of performance then WORM is the fastest followed by DVD. CD drives are relatively slow for both reading given the time to speed up a disc and also the writing of data.
- In terms of data volume the greatest capacity jukebox is still 12" WORM, followed by a large DVD jukeboxes.

DVD, CD-ROM and MO technology are all applicable regardless of whether the data is CAD drawings, maps, geophysical or relational database information. MO technology is heavily used in CAD/CAM and banking where data must be archived for over 25 years.

The mass storage of data using jukeboxes is a cost effective solution for data archiving when enormous quantities of on-line information must be searched and accessed very quickly.

Performance

CD, DVD and optical drives and jukebox systems are relatively slow devices compared to standard hard disks and RAID boxes. The swap time for even the fastest jukeboxes is a matter of seconds, whereas even the slowest hard disks have an access time of less than ten milliseconds.

Care should be taken when comparing the stated performance of the drives provided on data sheets against the performance of live systems. A number of factors will impact on performance not least of which is the time to write data across the network and also through the file system interface.

In this article we have tried to show that even using a sophisticated data management product like *Archimedia*, a jukebox system may appear relatively slow when compared to hard disk technology. The situation is complicated even further when a user tries to access the jukebox across a network, using networking protocols such as NFS.

The main performance problems are experienced in a loaded network environment when there are a number of read/write requests for different discs at the same time and there are more discs than drives. In this situation the requests will be queued and unless the software is very careful the jukebox will start to 'thrash'. This will be especially true when the layout of the data has not been optimized; say when the key information is stored on both sides of the same disc. The system may also appear to slow down if the cache is not large enough to hold many files and is being constantly flushed.

Archimedia is optimized to allow a large archive of data stored on optical discs in a jukebox to be accessed relatively quickly. However an optical library system should be viewed as complementary technology rather than as a replacement for hard disk storage or RAID.

Users wanting more detailed information about performance considerations or actual figures are requested to contact the technical development group at El Fresko Technologies.

Archimedia and Third Party Products

The Archimedia product has successfully been used with a wide variety of different applications including CAD/CAM, data archiving, Integrated Document Management and Workflow systems. Many of these products use relational database engines (such as ORACLE, Ingres and Sybase) to control the movement and integrity of the data.

Some products such as Helios Ether share make various technical assumptions about file systems and the Apple HFS protocol tries to 'sync' each file system every four seconds. This will cause a jukebox to thrash excessively unless careful measures are taken to partition data. In these environments it is better to mount only the directories on the hard disk and to explicitly manage the movement of data between the hard disk and the jukebox.

El Fresko Technologies is working with a number of system integrators in areas such as Pre-Press printing and CAD/CAM on the introduction of new systems which overcome these problems.

Contact El Fresko Technologies technical help desk for more details at support@elfresko.com.

Jukebox Control Software for Integrators

One key component of *Archimedia* is the Jukebox Control Software. El Fresko Technologies has licensed this software to a number of different application developers and integrators.

Jukebox Control Software (JBC) consists of a number of device drivers and command line utilities, as well as an Application Programmers Interface (API). It was developed with the needs of application developers and system integrators in mind and is a callable command library. JBC is used in applications such document management and workflow products, where developers need explicit control over the jukebox robotics and drives and know exactly where each file is located.

The software hides all complications of the jukebox robotics from the application/user and provides a transparent interface using standard tools. While it contains sophisticated queue management and error recovery software it is very straightforward to install and use.

Features and Benefits

Archimedia is a general-purpose jukebox management and archiving product aimed at the NT and Sun Solaris Operating Systems. It has a number of features that are crucial for obtaining optimal performance from a library system.

File based Cached Access	<i>Essential for fast access particularly for a slow device like a large CD jukebox or where many users are accessing files simultaneously</i>
Graphical Management Tools	<i>For ease of use and user friendliness, especially when installing and configuring the system (essential for the non technical system manager)</i>
Jukebox As One Large Disk	<i>Useful if file systems are larger than a single platter as it removes the need for users to write extra software to log where the file resides.</i>
Off-line Media Management	<i>Essential if jukebox starts to fill up</i>
Integrated Backup Facility	<i>Essential for fast disaster recovery</i>

Supported Environments

Host operating systems:

- Windows 2000 Server/Windows 2003 Serv
- Sun SPARC Solaris 2.6 and above
- Sun INTEL Solaris 2.7 and above

Clients:

- Any Windows 95/98/2000/XP/2003/Vista/7 client
- Any UNIX client

Summary

Archimedia has been widely acclaimed for its rich functionality and robustness. It is used in a variety of application areas such as printing, Pre-Press, Medical Imaging and Document Management as well as in Multi-Media and Imaging by some of the worlds largest corporations. By providing a general-purpose file interface (e.g. a drive letter) to a large archive device like a jukebox it makes the system integrator's task much easier. No special client side software is required for accessing the data; just reading and writing files using standard utilities.

The software is modular and can be easily tailored to meet most requirements such as multiple jukeboxes on the same server, lots of large files, millions of small files or tens of users accessing the same ten files.

Archimedia is a flexible and proven product. Don't take our word for it; talk to our users. Contact El Fresko Technologies for a list of references.