In this chapter you will learn to:

- document the storage and retrieval process in an information system
- describe the characteristics and operation of hardware devices used for storage and retrieval
- use a range of hardware devices and associated software to store and retrieve information and data
- store and retrieve data using a network
- compare different file formats for storing the same data, explaining the features and benefits of each
- use software features to secure stored data and information
- retrieve and use data in an ethical way

Which will make you more able to:

- describe the nature of information processes and information technology
- classify the functions and operations of information processes and information technology
- identify and describe the information processes within an information system
- recognise and explain the interdependence between each of the information processes
- identify and describe social and ethical issues
- describe the historical developments of information systems and relate these to current and emerging technologies
- select and ethically use computer based and non-computer based resources and tools to process information
- analyse and describe an identified need
- generate ideas, consider alternatives and develop solutions for a defined need
- recognise, apply and explain management and communication techniques used in individual and team-based project work
- use and justify technology to support individuals and teams

In this chapter you will learn about:

Storing and retrieving – the two-step process by which data or information can be saved and reloaded to allow for:

- other processing to take place
- a temporary halt in the system
- backup and recovery
- the transfer of data or information

Hardware for storing and retrieving

- hardware secondary storage devices, including:
  - magnetic disks
  - optical disks
  - network storages
  - flash memory
  - magnetic tapes
- the characteristics of hardware, including:
  - random or sequential access
  - volatile or non-volatile
  - permanent or non-permanent
- the trend to faster and greater storage capacity over time

Software in storing and retrieving

- hardware interface software
- file management software
- database management systems
- file formats for different data types
- Internet browser
  - used to access a machine independent data store
  - using search engines to access data
- encryption/password protection
- security of stored data whether stored centrally or distributed

Non-computer tools, including:

- paper based storage systems
- microfiche
- libraries

Social and ethical issues, including:

- the security of stored data
- unauthorised retrieval of data
- advances in storage and retrieval technologies and new uses such as data matching

Evaluation Copy www.pedc.com.au

6

TOOLS FOR INFORMATION PROCESSES: STORING AND RETRIEVING

Storing and retrieving is a two-part process; storing saves data or information and retrieving reloads data or information. Storing and retrieving supports all other information processes; it provides a mechanism for maintaining data and information prior to and after other information processes. The actual data or information is unchanged by storing and retrieving processes, rather the physical method of representing the data changes. For example, when saving data on a hard disk the storing information process physically represents the data using magnetic fields; when this data is later reloaded the retrieval process converts these magnetic fields into varying electrical signals that can be used by other hardware devices; in particular the CPU.

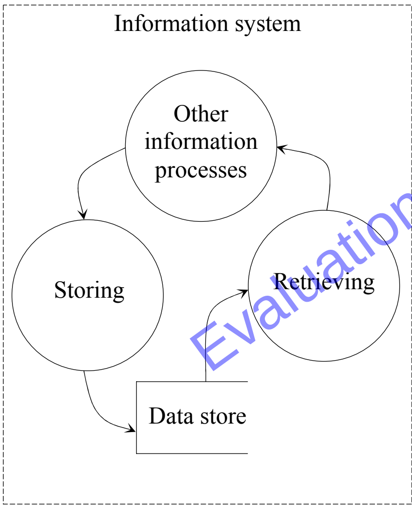


Fig 6.1
Storing saves data/information and retrieving reloads data/information.

The CPU can only process a limited amount of data at any one time; consequently it is necessary to maintain data and information both before and after processing. The CPU stores and retrieves data directly from primary storage; primarily RAM. However primary storage is volatile and non-permanent; to permanently store data requires secondary storage. As was discussed early in Chapter 5, data is retrieved from secondary storage into primary storage in preparation for processing by the CPU. Once the data has been processed it is returned to primary storage, and finally is stored on secondary storage.

Secondary storage is non-volatile; it does not require power, and is used to maintain a more permanent copy of the data or information. In this chapter we concentrate on the storage and retrieval of data to and from secondary storage; in particular we consider:

- the role of storing and retrieving within information systems,
- characteristics of storage hardware,
- the operation of common examples of secondary storage devices,
- software used for storing and retrieving,
- non-computer storage systems and
- social and ethical issues associated with storing and retrieving.

GROUP TASK Research
There are many different types of RAM available. Use the Internet to research and classify various types of RAM according to their data access speed, storage capacity and cost.

Evaluation Copy www.pedc.com.au

THE ROLE OF STORING AND RETRIEVING

Storing and retrieving is about preserving data, it allows data to be reused at a later time. What are the reasons for wishing to preserve data? Let us consider some common answers to this question.

TO ALLOW OTHER PROCESSING TO TAKE PLACE

Secondary storage has a much greater capacity than primary storage; furthermore it is not uncommon for files to be larger than the total capacity of primary storage. Consequently data, and even program instructions, must be retrieved from secondary storage as required by the process occurring at that particular time. Similarly once primary storage is full then data that has been processed is saved to secondary storage. In essence secondary storage is being used as an extension of primary storage; when such a system is formally implemented within an operating system the portion of secondary storage used as RAM is called virtual RAM. *Fig 6.2* shows a dialogue from Windows XP where the amount of virtual memory or virtual RAM is specified.

The situation becomes even more critical when various different processes are occurring at the same time; each process having its own data needs. For example, as I write these words my computer is running Microsoft Word, Internet Explorer, Microsoft Outlook, the Windows XP operating system together with various other software utilities for networking, virus detection, scanning, multimedia and faxing. Each of these processes uses data that is being swapped between secondary storage and primary storage as the need arises. It is the operating system's job to ensure each process is delivered the appropriate data and instructions at the required time.

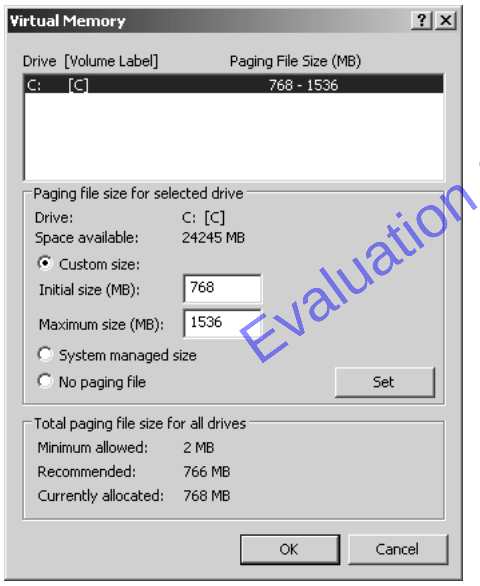
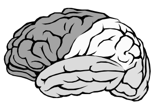


Fig 6.2
The amount of virtual memory or virtual RAM can be specified in Windows XP.



Consider the following:

The primary purpose of a file server is to store and retrieve data for a number of computers within a network. For this to occur all data must pass through the file server's primary storage (RAM) on its way to the network and then again on it's way back to the file server's secondary storage.

GROUP TASK Modelling and discussion
Construct a diagram to describe the flow of data in the above discussion. Why is it necessary for the data to move through the file server's RAM? Discuss.

GROUP TASK Discussion
File servers commonly receive simultaneous requests from different computers to either store or retrieve data. These requests appear to be processed simultaneously. How is this possible? Discuss.

Evaluation Copy www.pedc.com.au

TO ALLOW FOR A TEMPORARY HALT IN THE SYSTEM

It is uncommon for all information processes present in an information system to be completed in a single session. As a consequence provision must be made to halt the operation of the system for a period of time; obviously this requires all data to be permanently stored. A simple example would be a student completing an assignment. The assignment is unlikely to be completed in a single session; therefore the student saves the assignment, halts the system and then at some latter time reloads the saved data to complete the assignment. In this example, the collecting information process is interrupted for a period of time; this is commonly the case for most collecting processes. For example, an ordering function within an information system is activated each time a new order arrives, entering the order being a collecting information process. Between entering orders the computer is used for various other processes, hence orders must be stored to allow for a temporary halt in the ordering system. Furthermore, the collected data must be stored if it is to be used at a latter time by various analysing, processing and displaying information processes.

Consider the following:

The systems flowchart in *Fig 6.3* at right describes the logic and flow of data for an information system used to process the results of an assessment task.

It is not necessary to understand the meaning of each symbol on this diagram; systems flowcharts are not specified within the current IPT syllabus. However an understanding of the processing taking place is necessary. Firstly the teacher marks the assessment task. These marks are entered by hand into the teacher's mark book. At a later time the marks are entered into the computer where they are stored in the school database. At the same time the student names are being retrieved from the school database. Once all the marks have been entered they are scaled and stored in the school database. Finally a printout of the results is generated and students are given their results.

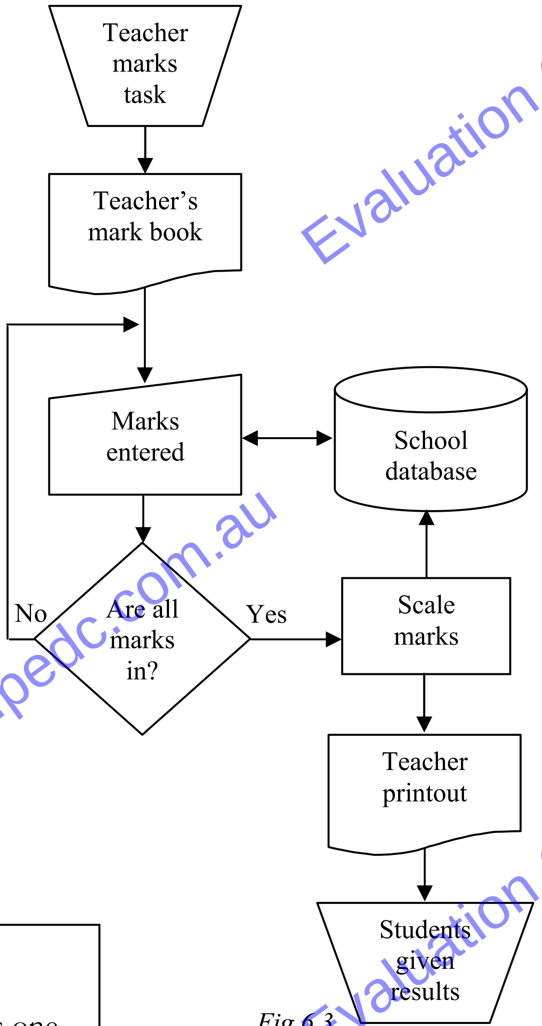


Fig 6.3
Systems flowchart for results from an assessment task.

GROUP TASK Classify
Classify each information process occurring in the above scenario as one of the 7 syllabus information processes.

GROUP TASK Identify and Discuss
Identify times within the above scenario where a halt in the system is possible. How does the ability to halt the system at these times assist the operation of the information system? Discuss.

Evaluation Copy www.pedc.com.au

BACKUP AND RECOVERY

Making a backup of data is the process of storing or copying the data to another permanent storage device, commonly recordable CD, magnetic tape or a second hard disk. Recovery of data is the opposite process where the data is retrieved or restored from the backup copy and placed back into the system.

> **Backup**
> To copy files to a separate secondary storage device as a precaution in case the first device fails or is lost.

The aim of creating backups is to prevent data loss in the unfortunate event that the original data is damaged or lost. Such damage most often results from hard disk failures; in fact it is inevitable that all hard disks will eventually fail. Some other reasons for data loss includes software faults, theft, fire, viruses, intentional malicious damage and even intentional changes that are later found to be incorrect. For backup copies to most effectively guard against such occurrences requires backups to be made regularly and that these backup copies be kept in a fireproof safe or at a separate physical location.

Even the most reliable computer will eventually break down and the consequences can be devastating if no backups have been made. Consider a small business with 100 clients; a total loss of data means loss of all client records, orders and invoices, together with any correspondence and marketing materials. Even if much of this information is maintained in paper-based storage the cost of recovering from such a loss is enormous compared to the minor costs involved to maintain regular backups.

There are two main types of backup that are commonly used; full backups and partial backups. A full backup includes all files whereas a partial backup includes only those files that have been created or altered since the last backup was made. Most operating systems include an archive bit stored with each file to simplify partial backups; each time a file is created or altered the archive bit is set to true. Backup and recovery utilities examine this bit to determine files to be included in each partial backup. Incremental partial backups set each archive bit to false once each file has been copied, whilst differential partial backups copy the files but leave the archive bit set to true. A common backup strategy involves completing a full backup (which sets all archive bits to false), followed by a series of partial backups. If a failure occurs then the full backup is restored first. If incremental backups were made then each must be restored in the order they were made. If differential backups were made then, once the full backup has been restored, only the most recent differential backup needs to be restored as it contains all changes since the last full backup.

The frequency at which backups are made depends on how critical the data is to the organisation. Commonly a full backup is made once a week with an incremental backup made daily. A further safeguard against data loss is to rotate the media used for backups; commonly three complete sets being used. This means that should one set of backups become corrupted then the previous set can be used for data recovery. In addition, maintaining different sets of backups means the system can be restored back to many different historical points in time. This is useful for restoring data that was inadvertently changed and for returning to a point prior to corruption occurring, such as before a virus attack.

> **GROUP TASK Research**
> Research and document the backup strategy used at your work or your school. How long do these backups take to produce and is it necessary to halt the system to perform each backup?

Evaluation Copy www.pedc.com.au

> **GROUP TASK Discussion**
> RAID is a system that uses multiple hard disks to store data. Should one disk fail then the other disks include sufficient data to not only rebuild the lost disk but to continue system operation. Do you think such a system removes the need to make regular backups? Discuss.

TO ASSIST THE TRANSFER OF DATA/INFORMATION

When we view a web page, receive an email or access data across a network we are, amongst other things, retrieving files from a storage device on a remote computer. The data or information on the remote computer must be stored before it can be retrieved and transferred to other computers. Furthermore, the data, once received by the local computer, must be stored locally prior to further processing and display. Hence the storing and retrieving information process is integral to the transmitting and receiving process.

There are software applications, in particular database applications, operating across networks where transferred data is stored within RAM on the receiving computer, however in general, most data received is stored locally as a file within secondary storage. For example, web browsers store a copy of every file retrieved from a website locally within a temporary Internet files folder on the hard disk; the browser retrieves these files from this folder prior to display. *Fig 6.4* shows the Parramatta Education Centre IPT page behind the contents of the temporary Internet files folder. In this screen shot the temporary Internet files folder was first cleared, hence each of the files shown is required to correctly view the web page shown. Similar storing and retrieving processes occur as an integral part of the transferring of most data across networks.

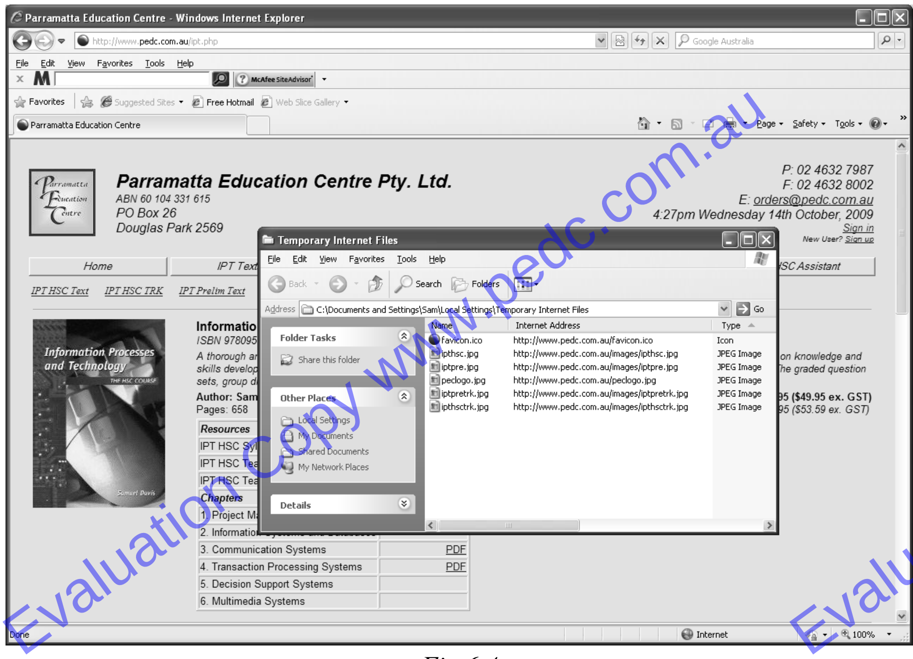


Fig 6.4
All files accessed using a web browser are transferred and stored locally.

> **GROUP TASK Discussion**
> When viewing a web page for the first time it often takes some time for all the images to appear, however on subsequent visits these same images appear virtually instantly. How can this be explained? Discuss.

Evaluation Copy www.pedc.com.au

HARDWARE IN STORING AND RETRIEVING

In this section we consider the characteristics and operation of a variety of commonly used secondary storage devices. Although our discussion is restricted to the operation of secondary storage devices it is important to remember that primary storage, such as RAM and ROM, is also hardware and that primary storage plays a vital role in the storing and retrieving information process.

GROUP TASK Discussion
RAM is certainly storage hardware, however it is integral to all seven of the information processes. Briefly discuss how RAM is used in each of the seven information processes.

Before we commence our examination of particular hardware storage devices let us discuss the meaning of some terminology commonly used to describe characteristics of such devices.

RANDOM OR SEQUENTIAL ACCESS

Random access refers to the ability to go to any data item in any order. Once the location of the required data is known then that data can be read or written directly without accessing or affecting any other data. The word random is used because the data can be accessed in any order, however in reality accessing any data item at random is unheard of; an equivalent, and perhaps more accurate term is direct access.

Sequential access means the data must be stored or retrieved in a linear sequence. For example, in *Fig 6.5* the sixth data item is needed so the preceding five data items must first be accessed. In terms of hardware devices, tape drives are the only widely used sequential storage devices. The time taken to locate data makes sequential storage unsuitable for most applications apart from backup.

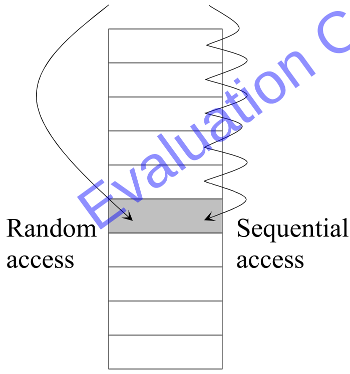


Fig 6.5
Random access versus sequential access.

Do not confuse random access files and sequential files with random access and sequential access storage devices. Random and sequential when used in relation to files describe the way software applications access files, when used in regard to storage devices these same terms relate to the physical storage of the data. Both types of file can be stored on either device; however a random access file stored on a sequential access device must physically be read and written sequentially, similarly a sequential file stored on a random access device despite being able to be physically accessed randomly will be read and written sequentially.

Consider the following:

Commodore released the Personal Electronic Transactor (PET) in 1977. The original PET used standard audio cassette tapes as its sole secondary storage medium, and had a massive 4K of RAM!

Fig 6.6
The Commodore PET 2001 released in 1977.

GROUP TASK Research
Use the Internet to research how data was stored on early personal computers that used audio cassette tapes.

Evaluation Copy www.pedc.com.au

VOLATILE OR NON-VOLATILE

Volatile computer storage requires a continuous electrical current to maintain data; if no electrical current is present then the data will be lost. On almost all computers RAM is volatile, if you do not save your data to secondary storage it is lost from RAM should a power failure occur. Dynamic RAM (DRAM) chips are particularly volatile; each storage area on these chips must be refreshed regularly to maintain their data whereas static RAM (SRAM) chips merely require electrical current to be present.

To reduce the effects of the volatility of RAM computers performing critical tasks are connected to uninterruptible power supplies (UPSs). The purpose of a UPS is to provide sufficient power to allow the contents of RAM to be written to secondary storage and then for the computer to be shutdown gracefully. At the time of writing non-volatile RAM chips had just been developed; currently such chips are used in specialised applications, however it is likely that eventually they will become part of all computer systems.

Predictably non-volatile storage does not require power to maintain stored data. Virtually all types of storage, apart from RAM, can be classified as non-volatile. Examples include ROM, magnetic disks and tapes, all types of optical storage and even flash memory.

PERMANENT OR NON-PERMANENT

No storage device is totally permanent; in reality there are only degrees of permanence. The meaning of the terms permanent and non-permanent largely depends upon the context in which they are used. Let us consider common uses of the terms permanent and non-permanent as they apply in different contexts.

Volatile memory such as RAM is certainly less permanent than any of the non-volatile forms of storage. Hence when comparing RAM with secondary storage it is common and appropriate to classify RAM as non-permanent and secondary storage as permanent.

When comparing different secondary storage devices permanence can be used to imply the inability to alter or erase data. Consider data stored on a hard disk, it can easily be altered or even erased, hence hard disks can be described as non-permanent. On the other hand the data on a non-recordable DVD or CD-ROM can be described as permanent; it cannot be altered or erased.

Another common use of the term permanent is in regard to archived copies of data. Commonly businesses make a complete copy of their financial records at the end of each financial year. This copy is placed into permanent storage, perhaps in a safe or even in a safety deposit box within their bank's safe. In this context it is not the medium on which the data is stored that determines permanence rather the term permanent describes the purpose of maintaining the secure copy.

A further common use is applied to backup copies of data, particularly in regard to networks. Files or complete storage devices that are included within regular backups are said to be permanent whereas data not included in such backups is said to be semi-permanent or even non-permanent.

GROUP TASK Discussion

It is often said that volatile storage devices are non-permanent and non-volatile storage devices are permanent. Do you agree? Discuss.

Evaluation Copy www.pedc.com.au

Evaluation Copy www.pedc.com.au

SET 6A

1. The CPU stores and retrieves data directly to and from:
   (A) secondary storage.
   (B) primary storage.
   (C) non-volatile storage.
   (D) permanent storage.

2. Virtual memory is used:
   (A) when there is insufficient secondary storage.
   (B) to remove the need to retrieve data from secondary storage.
   (C) when the amount of RAM is insufficient.
   (D) to speed up the processing of data.

3. What is the purpose of secondary storage?
   (A) To allow other processes to take place.
   (B) To allow for the system to be halted.
   (C) To assist the transfer of information.
   (D) All of the above.

4. Incremental backups are performed to:
   (A) ensure a complete copy of the data is maintained should a problem occur.
   (B) reduce the time used to perform backups.
   (C) ensure multiple backups are maintained.
   (D) secure data against unauthorised access.

5. Storage that requires power to maintain its contents is best described as:
   (A) volatile storage.
   (B) non-permanent storage.
   (C) non-volatile storage.
   (D) permanent storage.

6. The aim of creating backups is to:
   (A) prevent unauthorised access.
   (B) detect incorrect data.
   (C) protect against data loss.
   (D) remove the need for users to save their work.

7. When viewing a web page:
   (A) no data is stored locally, all the data remains on the remote web server.
   (B) all files used by the page are stored in RAM on the local computer.
   (C) each file needed to view the page is first stored locally in secondary storage.
   (D) the files within the page are sent directly to the display hardware.

8. Which of the following is NOT a characteristic of sequential storage?
   (A) To retrieve a data item requires retrieval of each preceding data item.
   (B) Individual data items can be retrieved from any physical part of the media without accessing any other data.
   (C) Tape is the only widely used sequential storage media.
   (D) For most applications sequential retrieval of data is slower than direct or random retrieval.

9. If the archive bit for a file is set to true then:
   (A) the file will be included an incremental backup but not in a full backup.
   (B) the file will be included in a full backup but not in an incremental backup.
   (C) the file will be included in both an incremental backup and a full backup.
   (D) the file will not be included in any backups.

10. Which of the following is NOT true of the storing and retrieving information process?
    (A) It supports all other information processes.
    (B) It alters the actual data or information.
    (C) It allows data to be reused
    (D) It maintains data prior to and after other processes.

11. Storing and retrieving assists each of the other information processes.
    Explain how storing and retrieving assists the
    - collecting,
    - organising, and
    - analysing information processes.

    Include examples within each of your explanations.

12. List and describe the differences between primary and secondary storage.

13. For each of the following, compare and contrast the meaning of the terms:
    (a) Volatile and non-volatile
    (b) Permanent and non-permanent
    (c) Random and sequential

14. Commonly commercial software is installed from CD-ROM. The installation involves various information processes.
    - List the sequence of information processes that would typically occur.
    - For each step in your sequence identify the hardware devices being used.

15. A small business receives on average 15 orders per day. These orders are processed as they are received using a commercial software package. The computer used to process the orders is also used for email, web access, word processing and various other administrative tasks.

    Recommend and justify an appropriate backup strategy.

OPERATION OF SECONDARY STORAGE HARDWARE

In this section we consider the operation and characteristics of magnetic storage, both disks and tape; optical storage such as various CD and DVD based technologies; network storage and finally flash memory technologies. Each of these secondary storage technologies is used to store and retrieve digital data in a non-volatile form.

MAGNETIC STORAGE

To understand the underlying operation of magnetic storage devices requires a basic knowledge of certain magnetic principles:

1. Magnets exert forces on each other known as magnetic fields. Such forces move from the north to the south pole of the magnet.
2. Magnetic fields are greatest at the poles.
3. Electrical currents produce magnetic fields.
4. There are only a few elements, primarily iron, cobalt and nickel, which can be magnetised. Materials that include these elements and that can be magnetised are known as ferromagnetic materials.
5. Different ferromagnetic materials behave differently when placed in a magnetic field.
   A. Some materials are easily magnetised by weak magnetic fields but when the field is turned off they quickly demagnetise; these materials are known as soft magnetic materials and are used during the process of storing or writing data.
   B. Some soft magnetic materials conduct electricity well when in the presence of a magnetic field but are poor electrical conductors when not. This phenomenon is called the magneto-resistance (MR) effect. MR materials are used during the process of retrieving or reading data.
   C. Some materials require a strong magnetic field to become magnetised however they retain their magnetisation when the magnetic field is turned off. These materials are known as hard magnetic materials and are used to produce permanent magnets. Such materials are the basis of magnetic storage media.

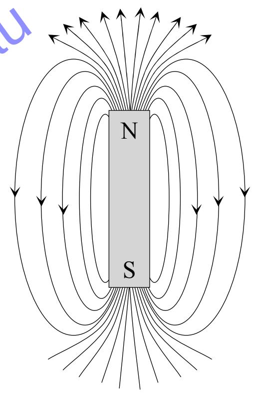


Fig 6.7
Magnetic forces move from north to south poles and are greatest at the poles.

To further assist our discussion let us first examine a microscopic detail of a typical piece of magnetic storage medium that already contains stored data (see *Fig 6.8* at right). This detail could be a section of a floppy disk, a hard disk platter or even a piece of magnetic tape; in each case hard magnetic material is used and the storage principles are the same.

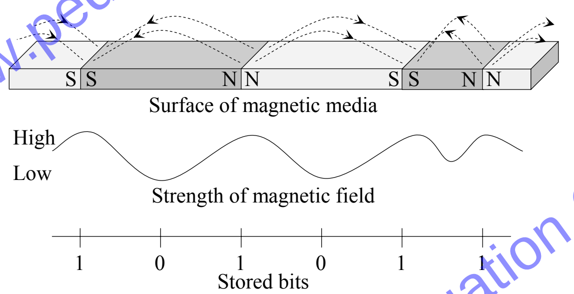


Fig 6.8
Microscopic detail of magnetic storage medium.

Digital data is composed of a sequence of binary digits, zeros and ones. These zeros and ones are equally spaced along the surface of the magnetic medium. High magnetic forces are present where the direction of the magnetic field changes; these points are really magnetic poles. It is the strength of the magnetic force that determines a one or a zero, not the direction of the magnetic force. Low magnetic forces occur between two poles and represent zeros. High magnetic forces are present at the poles and represent ones.

Evaluation Copy www.pedc.com.au

Evaluation Copy www.pedc.com.au

Consider the following:

At the time of writing (2009) the number of bits stored per inch (BPI) on the surface of a hard disk ranges up to around 1,000,000 BPI at the centre of each disk platter; this measure is commonly called linear density. This means a track on a hard disk can store some 40000 bits per millimetre. If *Fig 6.8* is the surface of a hard disk platter then the real width of the medium depicted would be approximately 1.5 ten thousandths of a millimetre; rather too small to print! Currently magnetic tape is available with a linear density of around 100,000 BPI resulting in some 4000 bits per millimetre.

FORUM

GROUP TASK Research

Investigate the linear density of various hard disks and magnetic tapes. During your research determine the relationship between linear density and areal density.

Storing or writing magnetic data

Magnetic data is written on to hard magnetic material using tiny electromagnets. These electromagnets form the write heads for all types of magnetic storage devices. Essentially an electromagnet is comprised of a copper coil of wire wrapped around soft magnetic material (see *Fig 6.9*). The soft magnetic material is in the shape of a loop that is not quite joined; this tiny gap in the loop is where the magnetic field is produced and the writing takes place.

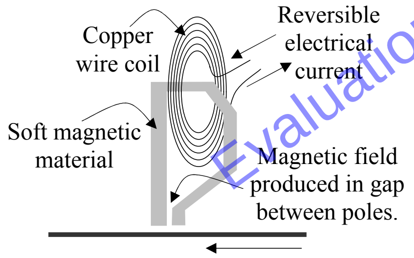


Fig 6.9
Detail of magnetic write head.

When an electrical current is present in the coil the enclosed soft magnetic material becomes magnetised, one end of the material becoming a north pole and the other a south pole. Hence a magnetic field is produced flowing from the north to the south. If the direction of the current through the coil is reversed then the direction of the magnetic field produced is also reversed. The magnetic field is strong enough for the hard magnetic material on the medium to be magnetised. A binary one is represented each time the direction of the magnetic field changes as a consequence of reversing the current into the coil. Zeros are represented when the direction of the current flow is constant and hence the direction of the magnetic field remains constant.

Retrieving or reading magnetic data

MR materials are the basis of most modern read heads; commonly this material contains around 80 percent nickel and 20 percent iron. Such materials are particularly sensitive to small changes in magnetic forces when a constant current is flowing through the material; that is they alter their resistance more noticeably. When stronger magnetic forces are detected, representing a 1, the current flow through the MR material increases and hence the voltage increases; similarly when the force is weaker the current and voltage decreases. These voltage fluctuations reflect the original binary data and are suitable for further processing by the computer.

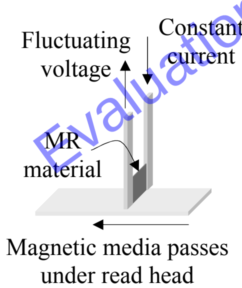


Fig 6.10
Detail of an MR read head.

MAGNETIC HARD DISKS

Hard disks store data magnetically on precision aluminium or glass platters. The platters have a layer of hard magnetic material (primarily composed of iron oxide) into which the magnetic data is stored. On top of this material is a layer of carbon and then a fine coating of lubricant. The carbon and lubricant layers improve the durability of the disk and slow down corrosion of the magnetic layer. Each platter is double sided, so two read/write heads are required for each platter contained within the drive's casing. At the time of writing most drives contain two to five double-sided platters requiring four to ten read/write heads. The casing is sealed to protect the platters and heads from dust and humidity.

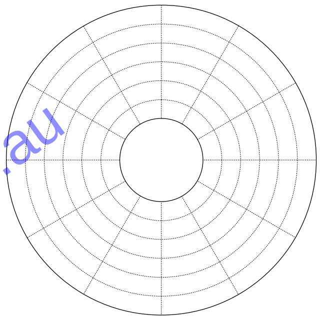


Fig 6.11
Each disk platter is arranged into tracks and sectors.

Data is arranged on each platter into tracks and sectors. The tracks are laid down as a series of concentric circles. At the time of writing a typical platter contains some one hundred thousand tracks with each track split into hundreds of sectors. The diagram in *Fig 6.11* implies an equal number of sectors per track; on old hard disks this was true however on newer hard disks this is not the case, rather the number of sectors increases as the radius of the tracks increase. Each sector stores the same amount of data, in most cases 512 bytes. The read/write heads store and retrieve data from complete sectors.

There are two motors within each hard drive; a spindle motor to spin the platters and an actuator assembly to move the read/write heads into position. The spindle motor operates at a constant speed; commonly from around 5,000 to 15,000 revolutions per minute. Whilst this is occurring the read/write head is moved in and out by the actuator assembly to locate the heads precisely over the required sectors on the disk platters.

Fig 6.12
Internal view of a hard disk drive.

Each read/write head is attached to a head arm with all the head arms attached to a single pivot point, consequently all the read/write heads move together. This means just a single read/write head on a single platter is actually operational at any instant. Each read/write head is extremely small, so small it is difficult to see with the naked eye. What is usually seen is the slider that houses the head. The air pressure created by the spinning platters causes the sliders to float a few nanometers (billionths of a metre) above the surface of the disk.

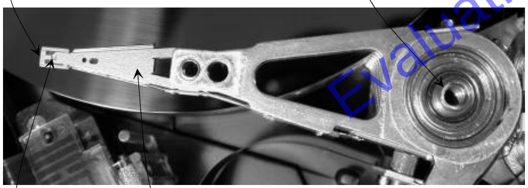


Fig 6.13
Expanded view of a head arm assembly.

Evaluation Copy www.pedc.com.au

Sophisticated circuits are required to control the accurate performance of the drive; in fact the processing power contained within a modern hard disk drive far exceeds the power of computers produced during the 1980s, furthermore they contain similar amounts of RAM in the form of cache. Hard drive circuits control the operation of the motors, communication with the CPU as well as checking on the accuracy of each read or write operation. Most hard disks contain their own built-in cache to significantly speed up access times. Data on sectors near the requested data is read into cache; commonly such data is subsequently required, consequently it can be accessed much faster from cache.

Fig 6.14
Underside of a hard disk drive showing the circuit board containing processing and cache chips.

Because the operation of a hard drive involves mechanical operations they will never reach the speeds possible with chip based storage technologies. Hard disks provide an economical means of permanently storing vast quantities of data. At the time of writing 250GB hard drives were common and drives exceeding 1TB were readily available. Currently, with the assistance of cache, hard drives are able to store and retrieve data at speeds exceeding 100MB per second.

Consider the following:

Older hard disk drives used the track (or cylinder) number, head number and sector number to determine the address of each sector (or block) of data. These addresses, known as CHS addresses, were translated via the computers BIOS (Basic Input Output System). Unfortunately such a system limited the size of hard disks to 1024 cylinders, 255 heads and 63 sectors per track equating to a capacity of 8.4GB.

As newer higher capacity hard drives became available and variable sectors were present on each track a new addressing system known as LBA (Logical Block Addressing) was introduced; this system essentially bypasses the computers BIOS altogether. LBA assigns each block (or sector) of data a unique sequential number; for example a drive with a total of 490,350,672 sectors would use LBA addresses from 0 to 490,350,671. The circuits within the hard drive translate the LBA address into the required physical address on the disks.

GROUP TASK Activity
Explain how 1024 cylinders, 255 heads and 63 sectors per track equates to a storage capacity of 8.4Gb?

GROUP TASK Research
Research specifications with regard to currently available hard disk drives. Determine the storage capacity, claimed data transfer rate, number of platters, total number of sectors and the storage size of each sector.

Evaluation Copy www.pedc.com.au

MAGNETIC TAPE

Magnetic tape has been used consistently for data storage since the early 1950s; the first such device being released commercially in 1952 by IBM (see *Fig 6.15*). At this early stage magnetic tape was the principal secondary storage technology; hard disk technologies first appeared in the late 1950s. The IBM 726 pictured featured six data tracks running parallel to the length of the tape, a seventh track was used for error checking. The linear density was around 100 bits per inch with a read/write speed of approximately 12,500 bits per second; current high performance magnetic tapes have linear densities exceeding 100,000 bits per inch and read/write speeds of more than 100 megabytes per second. These early devices where based on audio tape technologies; this has remained a common trend, many of today's tape drives borrow many of their components from audio or video tape drives.

Fig 6.15
The IBM 726 magnetic tape drive released in 1952.

Today magnetic tape is contained within cassettes or cartridges. Such cartridges range in size from roughly the size of matchbox to the size of a standard VHS tape. Tape is currently the most convenient and cost effective media for backup of large quantities of data. A single inexpensive magnetic tape can store the complete contents of virtually any hard disk; currently magnetic tapes (and tape drives) are available that can store up to 1TB of compressed data at only a few cents per gigabyte. The ability to backup the entire contents of a hard disk using just one tape far out way the disadvantages of sequential access; both backup and restore procedures are essentially sequential processes.

Fig 6.16
Various types of magnetic tape cartridges.

There are two different technologies currently used to store data on magnetic tape, helical and linear. Helical tape drives use technology originally developed for video and audio tapes; in fact the majority of the components, often including the actual tape cartridges, are borrowed directly from camcorders. Linear tape technologies were designed specifically for archiving data; hence in terms of data storage most linear systems perform their task more efficiently than helical systems.

Helical Technology

Helical systems contain a relatively large drum containing two pairs of read and write heads, each pair operates in isolation to the other. The tracks written by each pair of read/write heads cross each other at an angle of 40 degrees. Where two tracks intersect the magnetic forces combine. As both original forces are of equal strength, the direction of the combined magnetic force always remains closest to the direction of both the original forces and further away from either of their opposite forces In *Fig 6.17* the original forces are shown with open arrow heads and the combined force with a closed arrow head; the length of the arrows (or vectors) is an indication of the strength of each force. The result of these criss-cross tracks is a doubling of the linear density of the tape.

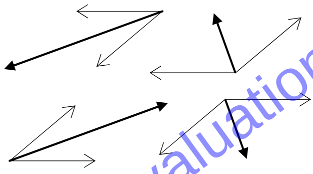


Fig 6.17
Original and combined force possibilities at the intersection of two tracks.

Pages 211 to 220

Not included in this Evaluation Copy

Evaluation Copy www.pedc.com.au

NETWORK STORAGE

Have you ever wondered how banks, government departments, web-based email systems, in fact any large computer network manages to store and retrieve data for many thousands of employees and customers? What's more, they manage to do this fast and securely. For example, consider EFTPOS, it is rare to have to wait more than a few seconds for a transaction to be approved, similarly logging into your hotmail account takes mere seconds. Furthermore, the large proportion of this time is attributed to the transmitting and receiving of the data rather than its storage and retrieval. Our aim in this section is to introduce some of the hardware used to perform high speed and secure network storage and retrieval processes. This includes not only providing data access to users of the system, but also creating backup copies of such vast quantities of data.

We shall consider the two most commonly used technologies: RAID (Redundant Array of Independent Disks) and also tape libraries. RAID provides fast data access combined with inbuilt fault tolerance. Tape libraries, as the name suggests, provide access to multiple magnetic tapes. Such libraries are primarily used for automated backup processes, however they also provide relatively fast retrieval of archived data.

RAID (Redundant Array of Independent Disks)

RAID utilises multiple hard disk drives together with a RAID controller. The RAID controller manages the data flowing between the hard disks and the attached computer; the attached computer just sees the RAID device as a normal single hard disk. The RAID controller can be a dedicated hardware device or it can be software running on a computer. In most cases the computer attached to the RAID device is a server on a network. This means a RAID device can be added to an existing network with minimal changes to existing server software and no changes to any other machines on the network.

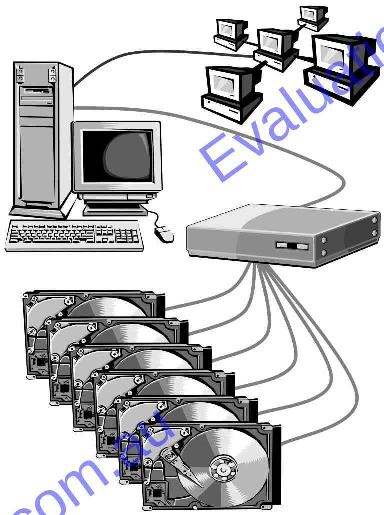


Fig 6.34
Main components of a RAID mass storage system attached to a network.

Simple RAID systems contain just two hard disks whilst large systems may contain many hundreds of disks. The RAID controller's job is to manage all these drives to improve data access speeds and fault tolerance. RAID is based on two basic processes, striping and mirroring. Striping improves read/write access times and mirroring improves fault tolerance and read times. Let us consider the operation of each of these processes.

Fig 6.35
A variety of different RAID devices.

Striping splits the data into chunks and stores chunks equally across a number of hard disks. During a typical storing or retrieving process a number of different hard drives are writing/reading different chunks of data simultaneously (see *Fig 6.36*). As the relatively slow physical processes within each drive occur in parallel, a significant improvement in data access times is achieved.

Evaluation Copy www.pedc.com.au

Mirroring involves writing the same data to more than one hard disk. *Fig 6.36* shows the simplest example of mirroring using just two hard disks where both disks contain identical data. When identical copies of data are present on different hard disks the system is said to have 100% data redundancy. Should one disk fail then no data is lost, furthermore the system can continue to operate without rebuilding data after the complete failure of a disk. Hence mirroring makes it possible to swap complete hard disks without halting the system; this is known as 'hot swapping'. Many larger RAID systems also include various other redundant components, such as power supplies; these components can also be 'hot swapped'. Data redundancy and the ability to 'hot swap' components improve the system's fault tolerance.

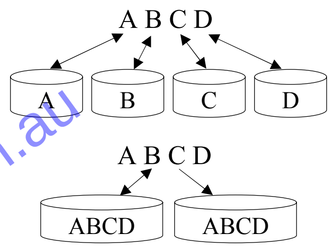
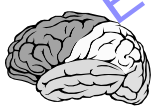


Fig 6.36
Striping (top) and mirroring (bottom) processes are the basis of RAID systems.

So mirroring improves the fault tolerance of the system, but what about read and write access times? Write access times are not reduced; in fact they may actually increase slightly due to the extra processing performed by the RAID controller. When mirroring, all data is written simultaneously to multiple hard disks; hence the time taken is similar to writing all the data to a single drive. On the other hand, retrieving data is quicker as any of the drives containing the data can be used; the RAID controller can make a choice, if the first drive is busy with another process the data can be retrieved from a different drive.



Consider the following:

In reality, the large majority of RAID systems utilise different combinations of striping and mirroring, known as RAID levels. RAID 0 uses just striping, RAID 1 just mirroring, all other RAID levels use a combination of striping and mirroring.

GROUP TASK Research
Use the Internet, or otherwise, to determine the most commonly used RAID levels. Describe how each of these RAID levels implements striping and mirroring.

Tape libraries

Have you ever made a complete backup copy of a hard disk? It involves manually swapping media and a good deal of time; these are major disincentives. Now imagine performing the same process for all the data held by a large organisation; hundreds or even thousands of tapes need to be swapped taking days or even weeks to complete. Clearly the backup process needs to be automated, this is the purpose of tape libraries.

Tape libraries, such as the one shown in *Fig 6.37*, include multiple tapes and multiple tape drives. A robotic system moves tapes between the storage racks and the tape drives. The tape drives are just normal single drives whose operation has been automated.

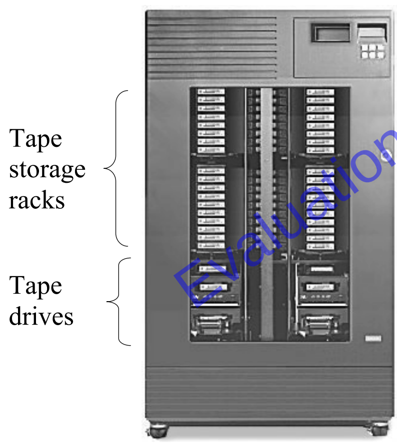
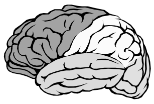


Fig 6.37
Qualstar's TLS-58132 tape library stores up to 340 terabytes of data.

Evaluation Copy www.pedc.com.au

Various different size tape library devices are available to suit the demands of different information systems. The smallest, such as Sony's TSL-SA400C in *Fig 6.38*, hold just four tapes and use a single drive; these devices provide capacities suited to most small businesses. Larger devices hold hundreds or even thousands of tapes and contain many drives. Large government departments and organisations link multiple tape library devices together; such systems hold hundreds of thousands of tapes and many thousands of tape drives. Backup processes on such large systems continue 24 hours a day, seven days a week.

Fig 6.38
Sony TSL-400C tape library.



Consider the following:

StorageTek's StreamLine™ SL8500 shown in *Fig 6.39*, has a minimum configuration of 1448 tapes which uses 64 tape drives. This modular system can be increased by combining up to 7 units to hold up to 70,000 tapes using 448 tape drives.

The system is capable of using tapes and tape drives of various types. Current tape capacities range from 20GB to 500GB per tape. Data storage speeds from 8 terabytes per hour up to 250 terabytes per hour are achievable depending on the type of tapes used, together with the configuration of the system

The average time taken for the robotics to place a tape in a drive is around 6.25 seconds. To identify individual tapes the robotic arms contain barcode readers, each tape being individually bar-coded. Furthermore redundant robotics, power supplies and electronics can be optionally installed to increase the fault tolerance of the system.

Clearly such systems are aimed at large corporate and government organisations that maintain extensive large-scale computer systems. Such systems are held in secure air-conditioned environments.

Fig 6.39
Exterior and interior of StorageTek's StreamLine™ SL8500 tape library.

GROUP TASK Activity
Determine the minimum and maximum storage capacity of the tape library system described above.

GROUP TASK Discussion
RAID and tape libraries help to secure data, but do they make it 100% secure? Discuss.

Evaluation Copy www.pedc.com.au

SET 6B

1. Which of the following best describes how binary data is represented on magnetic media?
   (A) A one is represented by a north pole and a zero by a south pole.
   (B) The direction of the magnetic field is used. One direction for ones and the other for zeros.
   (C) High magnetic forces represent ones and occur where the direction of the magnetic force changes. Low forces represent zeros.
   (D) A magnetic force exists where a one is represented and does not exist where zeros are represented.

2. Which of the following terms does NOT describe MR materials?
   (A) soft magnetic
   (B) conduct electricity better when close to a magnetic field.
   (C) used during the storing process.
   (D) used during the retrieving process.

3. The primary advantage of magnetic tape over other types of secondary storage is:
   (A) the speed of data access.
   (B) the ability to retrieve data sequentially.
   (C) that tape is much cheaper.
   (D) that tapes can be removed and stored off-site.

4. In a RAID device the process of striping is best described as:
   (A) Storing the same data on multiple drives.
   (B) Splitting up data, and storing each chunk simultaneously on different drives.
   (C) A technique for improving read times.
   (D) A method for improving fault tolerance.

5. The EFM coding system, together with merge bits, are used:
   (A) for error checking during the retrieval of data.
   (B) to restrict the length of both pits and lands so read errors do not occur.
   (C) to ensure both pits and lands are of sufficient length to be read accurately.
   (D) Both (B) and (C)

6. Helical tape systems:
   (A) use many components from audio and video tape drives.
   (B) write tracks at an angle to the length of the tape.
   (C) tend to wear out tape more rapidly than linear systems.
   (D) All of the above.

7. Which of the following best describes how binary data is represented on optical media?
   (A) Lands represent zeros and pits represent ones.
   (B) A change from land to pit or pit to land represents a zero whilst no change represents a one.
   (C) A change from land to pit or pit to land represents a one whilst no change represents a zero.
   (D) Changes in reflection are read as ones, whilst constant reflection is read as a zero.

8. Which of the following is true of sectors on hard disks?
   (A) Each track is always split into the same number of sectors.
   (B) All sectors on a particular hard disk store the same quantity of data.
   (C) The physical area of each sector is always the same.
   (D) Commonly the number of sectors per track increases as the radius of the track decreases.

9. The significant difference between CAV and CLV drives is:
   (A) Data passes under the read head of a CLV drive at a relatively constant speed; this is not the case with CAV drives.
   (B) The spindle motor operates at varying speed on a CLV drive but at a constant speed on a CAV drive.
   (C) CAV drives vibrate more as they spin at much greater speed than CLV drives.
   (D) The time taken to vary the speed of rotation in a CLV drive limits data transfer rates, hence CAV drives have higher data transfer rates.

10. Commonly the read/write head of an optical drive generates three laser beams. Why are three laser beams needed?
   (A) So that three data tracks can be read or written simultaneously.
   (B) One beam is used to read or write the data, whilst the others ensure the head remains centred on the data track.
   (C) One beam is used for the actual data and the other two are used for correcting errors within the data.
   (D) The use of three beams means that the laser does not need to be precisely focussed on the data track.

Evaluation Copy www.pedc.com.au

11. Describe the components and operation of the read/write head within a hard disk during:
    (a) a storing process.
    (b) a retrieval process.

12. Describe the components and operation of the read/write head within a CD-R drive during:
    (a) a storing process.
    (b) a retrieval process.

13. Describe how data is organised on the following storage media:
    (a) hard disks.
    (b) magnetic tape.
    (c) optical disks.
    (d) RAID devices.

14. Compare and contrast:
    (a) hard disk storage with magnetic tape storage.
    (b) recordable CDs and rewriteable CDs.
    (c) RAID devices and tape libraries.
    (d) Mirroring and striping used by RAID devices.
    (d) Flash memory with RAM.

15. Research both the storage capacity and data transfer rates for a variety of different models of RAID devices and tape libraries. Make up a table to summarise your results.

Evaluation Copy www.pedc.com.au

SOFTWARE IN STORING AND RETRIEVING

Software controls and directs the operation of all hardware, including all the various types of storage devices. Software causes hardware to perform processes that ultimately assist in achieving the system's purpose. So what software is used to perform storing and retrieving processes, and what does it do? To answer this question we first consider the various types of software operating behind the scenes to interface with storage hardware. We then consider the format of data files and how these formats affect storing and retrieving processes.

Virtually all application software utilises storing and retrieving processes, however there are particular types of software whose central purpose is managing the storing and retrieving of data. We consider examples of such software, namely file management software, database management systems and Internet or web browsers. Finally we discuss techniques for securing stored data, namely passwords and encryption of data.

THE HARDWARE TO SOFTWARE INTERFACE

It would be inefficient for every software application to direct and control all aspects of the storing and retrieving process. Rather such processes are split into various sub-processes performed by different programs, each piece of software being dedicated to a particular part of the storing and retrieving process. To identify and describe the software components involved in storing and retrieving let us consider a typical storing process; namely saving a file from within a software application to the hard disk. These steps could easily be adjusted to describe any storing process or reversed to describe a retrieving process:

1. Typically the user interacts with the application to initiate the save; this involves selecting a location for the file, and specifying a file name and storage format. This is a collecting information sub-process.
2. The application informs the operating system and passes it the location and name of the file. The operating system is now in control of the storing process.
3. The operating system directs the device driver associated with the appropriate storage device to proceed with the storing process. We discussed device drivers in Chapter 3 (see page 103); essentially the device driver provides a software interface between the operating system and the actual storage device.
4. Once the storage device is ready to commence the storing process it's device driver informs the operating system. The operating system then instructs the application to commence sending data directly to the device driver.
5. It is the job of the application to organise the data into the appropriate file format prior to sending it to the device driver. The device driver inturn passes the data from the application to the actual hardware storage device.
6. Within the storage device is further software, often called firmware, together with a buffer. The data stream progressively arrives and is held in the buffer as it waits its turn to be processed by the firmware. Firmware is permanently stored software within the hardware device; essentially the brain of the device.
7. The firmware controls the mechanics of the storage device to physically move components and store the data. It also reorganises the data as it leaves the buffer to suit the requirements of the actual storage device. For example, on a hard disk the data is split into appropriately sized chunks corresponding to the size of individual sectors on the disk. The firmware is not concerned with the file's format; it just sees the data as a stream of binary digits.

Throughout the whole process the operating system maintains ultimate control. Control messages are being relayed back and forth between the hardware and up through the various different software programs. These messages control the data transfer as well as ensuring the accuracy of the data.

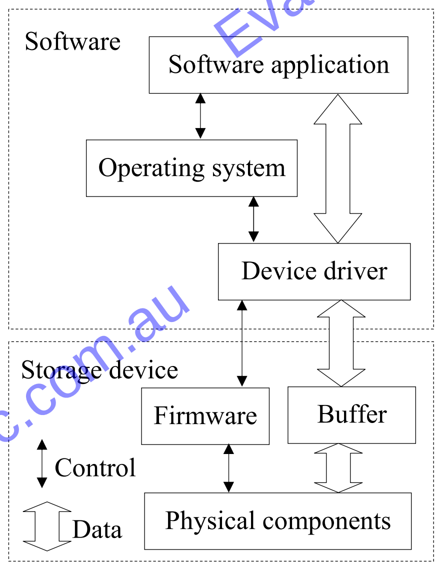
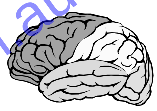


Fig 6.40
The interface between storage devices and software applications.

Fig 6.40 at right depicts the software and hardware components, together with arrows indicating the exchange of both data and control messages. This diagram applies to both storing and retrieving processes, hence the data arrows point in either direction. Essentially retrieving processes are the reverse of storing processes.

The dotted line around the software components in *Fig 6.40* also indicates software that executes on the main CPU; the firmware within the storage device being executed on a dedicated processor within the storage device. The gap between the two dotted rectangles represents the physical wires connecting the computer to the storage device. The arrangement of these wires, together with the connectors and rules for transferring data are part of an interface standard; the most common of these standards being ATA (Advanced Technology Attachment) and SCSI (Small Computer Standard Interface). The older parallel IDE (Integrated Drive Electronics) standards are often referred to as simply ATA, however the acronym PATA (Parallel ATA) is also used. The more recent Serial ATA (SATA) interface has largely replaced the Parallel ATA interface. Software with the ability to operate these interfaces is contained within the computer's BIOS (Basic Input Output System) and is loaded as part of the initial startup process.



Consider the following:

The operating system and the device drivers are stored on secondary storage. Our discussion above requires the operating system and the device driver stored on the hard disk to be loaded prior to retrieving data. It's a Catch-22 situation; you can't access the hard drive without the operating system and the device driver, yet to load this software requires the hard drive to be operational!

Fortunately, the hard disk contains firmware instructions and its own processor. Furthermore, the computer's BIOS is also firmware held on a dedicated chip. Both these firmware components are crucial to the successful start-up of all computers.

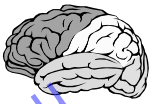
GROUP TASK Discussion
Obviously the operating system and device drivers do somehow get loaded from secondary storage. Discuss how this occurs.

GROUP TASK Investigation and discussion
Some firmware can be updated or edited whilst other firmware is completely permanent. Can you update the firmware within your home or school computer? Is firmware hardware or is it software? Discuss.

Evaluation Copy www.pedc.com.au

FILE FORMATS FOR DIFFERENT DATA TYPES

Organising data into a particular file format in preparation for storage is clearly an organising information process, however the file format chosen has implications in regard to the efficiency of storing and retrieving processes. The file format influences the size of the file and also the way in which the file may be retrieved.



Consider bitmap image files

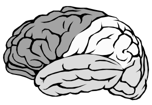
An image saved as a Windows bitmap (.BMP) uses significantly more storage than the same image saved as a JPEG file; the JPEG format includes the ability to significantly compress the data. Clearly storing and retrieving a compressed JPEG file takes less time than the larger BMP file. Furthermore, most files, including bitmap image files, are retrieved sequentially. As most bitmap files are arranged into rows of pixels commencing with the top (or bottom) row and ending with the bottom (or top) row the complete image cannot be displayed until the retrieval process is complete. Some bitmap formats, including JPEG, include the ability to arrange rows of pixels in non-sequential order such that a low-resolution version of the image is first displayed. As further rows are retrieved the resolution increases until eventually the complete image is displayed; JPEG files arranged in this manner are called progressive JPEGs.

GROUP TASK Calculate

A 505 by 391 pixel bitmap image has a colour depth of 24 bits. Saving this image as a Windows BMP requires 578KB of storage, however when saved as a JPEG the size is just 29KB. Explain how the BMP size can be calculated and express the compression of this BMP to JPEG as a ratio.

GROUP TASK Discussion

Compressed image file formats are used extensively on the web and also within digital cameras. Why is this? List and discuss reasons.



Consider the following:

Video data files are commonly organised in such a way that they can be progressively displayed as the file is being retrieved; this process is known as streaming. On many websites it is possible to jump to a later scene without the need to download all preceding scenes. On other websites the user must wait for all intermediate scenes to be retrieved before the later scene is displayed.

GROUP TASK Research

How can the differences described above be explained? Discuss.

GROUP TASK Discussion

Database Management Systems are able to retrieve specific records within a single file. How is this ability similar to the ability to jump directly to scenes within a video file? Discuss.

Evaluation Copy www.pedc.com.au

FILE MANAGEMENT SOFTWARE

File management software is used to logically organise files on secondary storage devices. Most operating systems include file management software. For example, Windows Explorer is an integral part of the Windows family of operating systems. Such software is not concerned with the data within files but rather with the manipulation of complete files. The aim is to present a logical arrangement of the files to the user and to provide processes for manipulating files within this arrangement.

What do we mean by the term logical when referring to the arrangement of files? Different storage devices physically store data differently, however file management systems are able to translate this physical arrangement of data into a consistent logical arrangement. When using file management software the directory structure appears similar regardless of the type of storage device. For example, the screen from Windows Explorer shown in *Fig 6.41* includes hard disks, removable storage and network drives, yet all are presented to the user in a similar consistent manner. Furthermore, from the user's perspective, opening any of these devices and manipulating files is performed using identical actions. In effect file management software hides the physical details of where and how files are stored and manipulated.

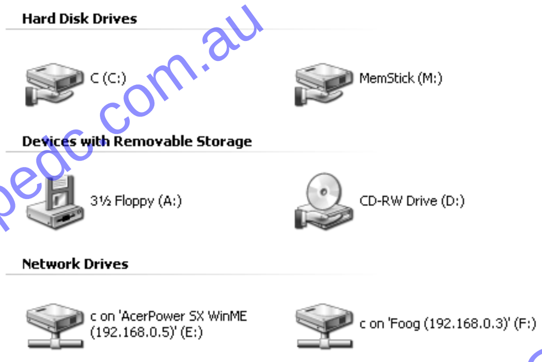


Fig 6.41
Various types of storage device are accessed using the same user interface.

So what is this logical arrangement? Files are arranged into a hierarchical structure of directories or folders. Each storage device has a root directory, which may contain both files and other directories; each of these directories also contains files and/or further directories and so on. For example, in *Fig 6.42* the file abc.gif is within the directory called Images, which is within the directory called IPT Text. Actually a directory is merely another file, it contains the name, location and various other details of each of its files. Recall our discussion in regard to archive bits on page 192; there is a similar bit set for files that are directories. The file management system, in consultation with the operating system, reads this bit to determine files that are directories.

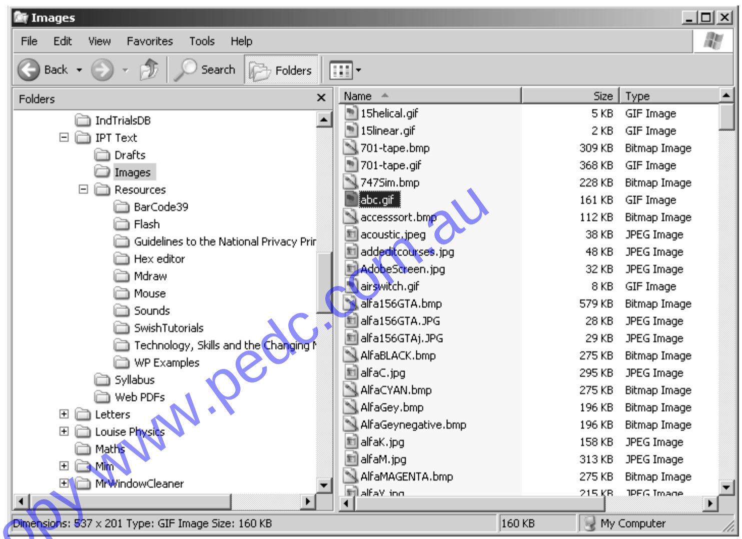
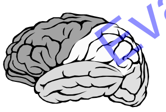


Fig 6.42
Screen shot from Windows Explorer included with Microsoft's Windows XP operating system.

GROUP TASK Activity

Most file management software includes new, cut, copy, paste and rename functions. Use each of these functions to manipulate files on either your home or school computer. List and describe other functions available.

Evaluation Copy www.pedc.com.au



Consider the following:

Data on most current hard disks is physically stored in individual 512-byte sectors. Many operating systems utilise a storage system known as FAT, which combines multiple sectors into clusters (typically from 4 to 64 sectors per cluster); each file resides within a particular number of complete clusters. A file allocation table (FAT) on the disk contains an entry for every cluster. These entries indicate whether a cluster is free, damaged or being used to store part of a file. If it is being used then this entry either points to the next cluster holding data for the file or it contains a flag indicating it is the last cluster for the file.

The directory file contains entries for each file within the directory. Each of these entries includes the address of the first cluster on the disk containing the file. When the operating system wishes to access a file it retrieves the address of the first cluster from the directory file; subsequent cluster addresses being obtained from the FAT. These addresses are submitted to the hard disk, which responds by retrieving the data within the sectors corresponding to the specified cluster addresses.

GROUP TASK Discussion
Deleting a file does not actually remove the data. Based on the above information, discuss what is likely to be occurring during a typical delete operation.

GROUP TASK Discussion
A file containing exactly 30000 bytes is being stored. Assuming each cluster contains four 512-byte sectors, calculate the number of clusters used and describe the changes made within the file allocation table.

DATABASE MANAGEMENT SYSTEMS (DBMS)

Database management systems are software applications used to store and retrieve data within databases. Most databases contain various types of data arranged into multiple tables where each table is composed of records. We discussed the organisation of such data back in Chapter 4 (p154-155); it may be worthwhile reviewing these pages. Now imagine the size of a database maintained by even a small organisation or business; it is likely to contain many tables and many thousands of records and furthermore many users have simultaneous access to this data. For example, a user can be entering an order for a customer whilst another user is analysing sales trends; they are both accessing the same data. In large organisations the number of users and the number of records becomes massive, perhaps many thousands of users and many millions of records. It is the job of the DBMS to manage the storing and retrieval of this data so that all users have access in a logical and efficient manner. Clearly DBMS software does not simply retrieve complete files.

GROUP TASK Calculate
Most schools maintain their timetable in a database. Details of each student being held in one table, details of each class in another, and a further table linking each student to their classes. Calculate the approximate number of records held in each of these tables within your school's timetable database.

Evaluation Copy www.pedc.com.au

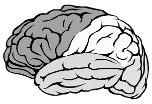
Pages 231 to 240

Not included in this Evaluation Copy

Evaluation Copy www.pedc.com.au

SOCIAL AND ETHICAL ISSUES ASSOCIATED WITH STORING AND RETRIEVING

Social and ethical issues in regard to the storing and retrieving of data are largely concerned with ensuring stored data can only be accessed and used by authorised persons for authorised purposes. Previously in this chapter we examined the use of passwords and encryption techniques, such tools are effective in terms of preventing unauthorised access to data, however they do not protect the data against unauthorised use by authorised users. In Chapter 1 (p19-20) we discussed the security of data and information, and in particular some of the strategies used to address security concerns. In this section we consider examples of particular social and ethical issues arising as a consequence of the storage of data; to assist your discussion of these issues it would be worthwhile reviewing page 19 and in particular page 20.

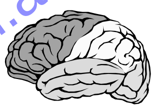


Consider the following:

Many taxation office employees have access to individual's taxation records; such access is necessary for the completion of their duties. As a consequence it is possible, and perhaps likely, that some of these employees will access and read their friend's tax returns. Such events are difficult to prevent; privacy laws are a deterrent but in this case somewhat difficult to enforce. The breach must first be detected; to do this requires costly and constant monitoring of user's access to individual records. Furthermore, such detailed monitoring of employees raises further ethical issues in regard to privacy.

GROUP TASK Discussion

Imagine you are employed by the tax office. Would you be tempted to read your friends' tax returns? How would you feel about being constantly monitored? Discuss.



Consider the following:

A shop owner would not leave their cash register full of cash and the front door open at the close of business, however many businesses effectively do this with their sensitive data and information. They simply do not recognise the risks and possible effects of unauthorised access to such data.

Furthermore, on the whole even large businesses are unable to detect unauthorised access has even occurred let alone be able to identify the perpetrator. History tells us that all security measures are eventually circumvented; hence regardless of the security systems in place no data is ever truly secure.

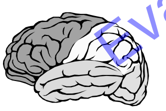
GROUP TASK Discussion

Describe possible effects of unauthorised access to sensitive business data.

GROUP TASK Research

Do you think secure public key encryption systems will ever be broken? Use the Internet to gather various opinions to assist your response.

Evaluation Copy www.pedc.com.au



Consider the following:

In 1990 the federal government approved legislation authorising the use of a system known as "The Parallel Data Matching Program". This legislation was required to override various provisions existing within the *Privacy Act 1988*, in particular to legalise the use and linking of personal data as part of the data matching process. The Data Matching Agency (DMA) was subsequently created to implement the system under the control of Centrelink. The DMA uses data sourced from various government departments and agencies including the Departments of:

- Social Security (DSS)
- Veterans' Affairs (DVA)
- Employment, Education, Training and Youth Affairs (DEETYA)
- Health and Family Services
- the Australian Taxation Office (ATO)
- Centrelink

The data matching process links the individual personal records held by each of these departments in an attempt to identify various fraudulent and illegal activities. In many cases, tax file numbers are used, however it is also common for names, addresses and other private data to be the basis of the data matching process.

The purpose of the DMA is to detect:

- instances of tax evasion
- fictitious or assumed identities
- incorrect payments from support agencies
- inaccurate income disclosures

The DMA has access to and processes data on virtually every single individual resident of Australia; hence initially everyone is a suspect. Should inconsistencies be determined then the presumption is that the person is guilty. This is the opposite of other investigative procedures whereby evidence points to particular individuals who are then investigated in an attempt to gather further evidence.

GROUP TASK Discussion
Do you think the fraudulent activities detected by the DMA justify its extensive use of personal information? Discuss

GROUP TASK Discussion
The Privacy Commissioner has described the above data matching process as "the information society's equivalent of drift-net fishing". What do you think the commissioner meant by this statement? Discuss.

GROUP TASK Discussion
Data matching does not merely look for perfect matches; it also links records that have a certain level of similarity. Obviously such matches are often incorrect. List and describe possible consequences of such incorrect matches being assumed to be accurate.

Evaluation Copy www.pedc.com.au

HSC style question:

(a) Describe the movement of data and the restrictions on the speed of data access between secondary storage, primary storage and the CPU during a process that analyses large amounts of data.

(b) Explain how binary digits are represented on magnetic tape.

(c) Outline reasons why most organisations still maintain paper-based filing systems in addition to their computer-based storage systems.

(d) Describe TWO techniques that aim to secure digital data so it cannot be read by unauthorised users.

Suggested Solutions

(a) During analysis data is retrieved by the CPU from RAM (primary storage). If the data is not present in RAM then it must be retrieved from secondary storage (usually a hard disk) into RAM. The CPU operates much faster than RAM and RAM operates much faster than secondary storage. Because large amounts of data are being analysed then RAM cannot be filled quickly enough from the slower secondary storage, so RAM cannot keep up with the demands of the CPU. The analysis will only operate at the speed of the slowest device – secondary storage in this case.

(b) The bits are equally spaced along a track on the surface of the magnetic tape. When the direction of the magnetic field changes the magnetic force is greatest – such points represent binary ones. Binary zeros are represented where the magnetic field does not change direction and hence the force is lower.

(c) Possible reasons organisations maintain paper-based filing systems include:

- The existing computer system does include the functionality required to store all the data used by the organisation and it is not cost effective to update to a computer system that can perform these functions.
- The paper records are not required by other information processes, therefore there is no need for them to be digitised.
- The original of many documents must be kept for legal reasons. For example, original signatures and seals placed by courts cannot be reproduced digitally.
- The organisation does not own and cannot justify purchasing the hardware to digitise their paper records.
- Some data is not suited to computer-based storage. For example, hand written notes, instruction manuals, cash register receipts, etc.

(d) Passwords can be used so that the system can identify that a user is who they say they are. Permission to read data being based on the user name.

Encryption involves using an algorithm to scramble the data using a key. The key must be known during the decryption process. Therefore people who do not have the key see scrambled data.

Evaluation Copy www.pedc.com.au

CHAPTER 6 REVIEW

1. Optical media includes:
   (A) Hard disks and tape cartridges.
   (B) Hard disk drives and tape drives.
   (C) CD drives and DVD drives.
   (D) CDs and DVDs.

2. Flash memory is solid state and non-volatile, this means:
   (A) it is portable and is difficult to destroy.
   (B) power is required to maintain the data, but no mechanical parts are used.
   (C) it contains no moving parts, and requires no power to maintain its contents..
   (D) it is contained on a microchip and does not require power for data storage.

3. Microfiche stores data:
   (A) magnetically
   (B) photographically
   (C) optically
   (D) electrically

4. Data is stored on a single continuous track on all:
   (A) CDs.
   (B) DVDs.
   (C) magnetic tapes.
   (D) hard disks.

5. Electromagnets produce magnetic forces when power is applied, they are used during:
   (A) optical storing processes.
   (B) optical retrieving processes.
   (C) magnetic storing processes.
   (D) magnetic retrieval processes.

6. The process of linking records from multiple data sources is known as:
   (A) data retrieval.
   (B) data matching.
   (C) record linking.
   (D) drift-net fishing.

7. Drives capable of storing data on rewriteable optical media:
   (A) have a laser capable of operating at two different intensities.
   (B) contain MR material within their read/write head.
   (C) contain lasers capable of operating at three levels of intensity.
   (D) produce significant levels of vibration that commonly cause read and write errors.

8. Software that assists the user to copy, delete and paste complete files is known as:
   (A) a database management system.
   (B) a tape library.
   (C) an operating system.
   (D) file management software.

9. If a collection of data will only ever be encrypted and decrypted by a single machine or user then:
   (A) a password is sufficient security.
   (B) single key encryption is suitable.
   (C) public key encryption should used.
   (D) All of the above should be used.

10. The read/write heads in a linear tape drive commonly have each write head positioned between a pair of read heads. Why is this?
   (A) So the tape can be maintained in the correct vertical position.
   (B) To enable data to be read, written and then reread without the need to rewind.
   (C) So data can be written, then verified in either direction.
   (D) It is cheaper to produce, as such components are part of domestic camcorders.

11. List and describe the main components of each of the following devices:
   (a) hard disk drive.
   (b) DVD drive.
   (c) RAID device.
   (d) tape library.

Evaluation Copy www.pedc.com.au

12. Discuss each of the following:
    (a) How can flash memory cards be used to help secure data?
    (b) What are the differences between helical tape and linear tape systems?
    (c) RAID devices help to protect data against various problems. What are these problems?
    (d) Tape libraries use multiple small magnetic tapes, in some cases many thousands of them. Doesn't it make more sense to just use much larger tapes?

13 For each of the following scenarios:
- Identify and describe any social and ethical issues arising.
- Suggest a method for securing the data to prevent such issues arising in the future.

    (a) The hard disk on a file server fails. This results in many employees not being able to work for a total of five days whilst a new disk is installed, all the software loaded and configured and finally the data is restored from backups. Most employees are not particularly concerned however management subsequently fires the entire IT department.
    (b) A mail order business commences trading over the Internet. Unfortunately they begin receiving complaints from customers that their credit card details are being used to purchase goods from all over the world.
    (c) Various employees, via casual chitchat, form the opinion that their private and sensitive business emails are being read by at least one of the company's network administrators. There suspicions are shared with management, who respond by developing a code of conduct that includes a statement discouraging such activity. However, a sub-clause is included permitting senior management to read any emails as they see fit.

14 During the storing and retrieving process the actual data is unchanged, however its physical representation changes and so too does the method of binary representation. In essence the raw data is being reorganised various times as part of storing and retrieving processes.

Identify and describe each reorganisation of data that occurs during the process of:
    (a) saving a file to a hard disk..
    (b) retrieving a file from a CD-ROM.
    (c) saving a file to a RAID device.

15 Storage devices are composed of various sub-systems that are ultimately composed of individual hardware and software components. Each component possesses characteristics that make it suitable for its particular task.

For each of the following components, describe:
- how the component is used by the device during storing and/or retrieval of data.
- characteristics of the component that makes it suitable for the task it performs.

    (a) Electromagnets
    (b) Lasers
    (c) Spindle motors
    (d) Opto-electrical cells
    (e) The dye layer within a CD-R.
    (f) The crystalline layer within a CD-RW.

Evaluation Copy www.pedc.com.au