Syllabus outcomes

5.3.2 Acquires and manipulates data and information in an ethical manner.

Overview

The emphasis in this chapter is on data handling. You will learn about data and information, data coding, data sources and data types. The storage of data, both primary and secondary, is examined. The chapter concludes by outlining data transmission, data compression and data security.
3.1 Data and information

Data is the raw material entered into a computer system. This raw material could be text, numbers, graphics, audio, animation or video. Data can be thought of as unorganised facts that mean little by themselves, such as the letters typed into a word processor. Information is data that has been ordered and given some meaning by people, such as a word processed document that we can read and understand.

There are many different ways of converting data into information, such as combining data, displaying data in an understandable way or deleting irrelevant data. Information depends on the data entered. For example, data might be measurements, survey results, a document, or photograph. By itself, this data would not mean much, but when it is transformed in an appropriate way it can provide useful information. This distinction between data and information is important. It explains why people can collect huge amounts of data but not satisfy their information needs.

Information is valuable if it is relevant, accurate, current and appropriately presented. Information must be attractive and easy to use. An appropriate format needs to be chosen. This often depends on the users of the information and how the information will be presented. Some people will understand information more completely if it is presented graphically, while others will understand it better in a table. Information is displayed using a range of hardware devices such as a screen, printer, or speaker. Data is represented in analog or digital form.

Analog

Analog data is represented by using continuous variable physical quantities such as voltages. For example, an analog clock shows a second hand continually sweeping around the clock face, measuring small units of time. At any moment the clock will give you an instant measure of the exact time, since you can estimate parts of a second as the second hand moves. Most natural events in the real world, such as temperature, light and pressure change smoothly and slowly like the hands of the clock (see Figure 3.1).

Sounds, images and video are naturally in
analog form. Analog signals are pulses, usually electrical or optical, in the form of continuous waves.

**Digital**

*Digital* data is represented using discrete measurements in the form of digits or numbers. For example, a digital clock shows the time as a certain number of hours and minutes. Here the time does not change continuously, but in a series of steps, jumping from second to second and minute to minute. Data is not represented continuously, but in discrete quantities using digits. Numbers, text and other characters are naturally in digital form.

Information technology uses data in digital form using only two digits: 0 and 1. Two digits are easily represented electronically by circuits in the computer being off or on. The digit 0 is used to represent the electronic state of ‘off’ and the digit 1 is used to represent the electronic state of ‘on’. Each on or off digit is called a bit (Binary digit). A *bit* is the smallest unit of data stored in a computer. A group of eight bits is called a byte (see Section 3.2 for more information).

Data is processed in words. A *word* (or word size) is the number of bits processed by the computer at one time. When the computer moves bits from one place to another it takes one word at a time. Most computers use word sizes of 8, 16, 32 or 64 bits. The bigger the word size, the faster the computer.

**Exercise 3.1**

1. True or false?
   - a. Information must be attractive and easy to use.
   - b. Sounds, images and video are naturally in digital form.
   - c. Information technology works with data in analog form.
   - d. A bit is a binary digit.
   - e. A group of eight bytes is called a bit.

2. Copy and complete the following sentences:
   - a. Data can be thought of as _____ facts that mean little by themselves.
   - b. Information is _____ using a range of hardware devices.
   - c. _____ data is not represented continuously, but in a series of steps.
   - d. A _____ is the number of bits processed by the computer at one time.
3.2 Data coding

The binary number system is used to represent digital data in a computer system.

Decimal and binary

The decimal system is an arithmetic system using a base of ten. It consists of ten digits (0 to 9). The decimal system uses place value headings of 1 (10^0), 10 (10^1), 100 (10^2), 1000 (10^3), and so on. The position of the digit determines the worth of a digit. For example, the decimal number 6543 represents 6 × 1000 + 5 × 100 + 4 × 10 + 3 × 1. The binary system uses place value like the decimal system.

The binary system (or binary code) is an arithmetic system using a base of two. It consists of two digits, 0 and 1. The decimal system uses powers of ten for its place value and the binary system uses powers of two. The place value of a binary number starts at 1 (or 2^0), the next heading is a 2 (or 2^1), then a 4 (or 2^2) and
The word ‘byte’ is an abbreviation from the words ‘BinarY digit Eight’.

Data handling continues to double in value for each additional place value. These place headings continue indefinitely like the decimal system. However, data in a computer is usually grouped using only 8 place value headings or 8 bits (0s or 1s)—and, as noted above, a group of 8 bits is called a byte and it represents a single unit.

To change a binary number into a decimal number, add the appropriate powers of two or place values (128, 64, 32, 16, 8, 4, 2, 1).

Example: Convert 01001110\text{\scriptsize{2}} into a decimal number (the subscript indicates the number is in base two):

<table>
<thead>
<tr>
<th>Powers of two Value</th>
<th>2^7</th>
<th>2^6</th>
<th>2^5</th>
<th>2^4</th>
<th>2^3</th>
<th>2^2</th>
<th>2^1</th>
<th>2^0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary number</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

01001110\text{\scriptsize{2}} = 0 \times 128 + 1 \times 64 + 0 \times 32 + 0 \times 16 
+ 1 \times 8 + 1 \times 4 + 1 \times 2 + 0 \times 1 
= 64 + 8 + 4 + 2 
= 78\text{\scriptsize{10}}

Binary number 01001110\text{\scriptsize{2}} equals 78, as a decimal number.

To change a decimal number into a binary number, divide the place values for binary (128, 64, 32, 16, 8, 4, 2, and 1) into the decimal number. The result of the division is the binary digit, and the remainder is divided by the next place value. This process is repeated for all place values.

Example: Convert 109 into a binary number:

<table>
<thead>
<tr>
<th>Powers of two Value</th>
<th>2^7</th>
<th>2^6</th>
<th>2^5</th>
<th>2^4</th>
<th>2^3</th>
<th>2^2</th>
<th>2^1</th>
<th>2^0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary number</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

109\text{\scriptsize{10}} = 0 \times 128 + 1 \times 64 + 1 \times 32 + 0 \times 16 
+ 1 \times 8 + 1 \times 4 + 0 \times 2 + 1 \times 1 
= 64 + 32 + 8 + 4 + 1 
= 01101101\text{\scriptsize{2}}

Decimal number 109 equals 01101101\text{\scriptsize{2}} in binary code.

Many calculations are performed on the binary code. Addition using the binary system is the same as addition in the decimal system. It is important to keep the numbers under each other to maintain their correct place value. Start by adding the units column and remember that 1 plus 1 in binary is 10.

Example: Add 01101101\text{\scriptsize{2}} and 00100110\text{\scriptsize{2}}:

\begin{align*}
01101101 \\
00100110 \\
\text{\scriptsize{11010011}}
\end{align*}
**ASCII**

ASCII (pronounced ‘ass-kee’) is a code that represents letters, numbers, and symbols as binary code. It is the standard method used by personal computers and stands for the American Standard Code for Information Interchange. A standard 7-bit ASCII was designed when computers were not extensively used outside of the US and UK. It only allows 128 different characters (2⁷). This was not large enough to handle all the characters used in the languages of western Europe.

Extended ASCII was developed using an 8-bit binary code to extend the character set. It is now the most widely used code. The 8-bit ASCII code allows for 256 different characters (2⁸), starting with 00000000 and going through to 11111111. There are 96 keyboard characters including 26 upper case letters, 26 lower case letters, 10 digits, and 34 symbols. The 8-bit ASCII code also contains 32 control characters that are used to control computer functions such as a ‘carriage return’ and a ‘form feed’.

To change a character into its ASCII binary code you need to refer to the ASCII table in the appendix. For example, the letter ‘K’ is the decimal number 75 and its ASCII binary code is 01001011. The computer uses the binary number 01001011 to represent the letter ‘K’.

It should be remembered that computers work with other data types besides text. However, a group of bits can also represent every other data type such as graphic, audio and video. We will examine data coding for these data types in later chapters.

**Bits and bytes**

A bit is the smallest unit of data stored in a computer. A byte is a group of 8 bits that represents a single unit. If the data is text then a byte would represent a character, such as a letter, a number, a punctuation mark or a space. The prefixes kilo, mega, giga, and tera are then added and more commonly used to measure data storage (see Table 3.1). The lowercase ‘b’ is used to represent a bit while the uppercase ‘B’ is used to represent a byte. However, it is common in speech for the byte to be omitted so that a gigabyte is simply a ‘gig’ or G.
Exercise 3.2

1. What am I?
   a. An arithmetic system using a base of two.
   b. The ASCII code for ‘full stop’ character.
   c. A code that changes letters, numbers and symbols into an 8-bit binary code.
   d. Another name for the binary system.

2. True or false?
   a. The place value in the binary system doubles in value for each additional place.
   b. An 8-bit ASCII code can represent at most 255 different characters.
   c. The binary number 10 represents the same amount as the decimal number 10.
   d. ASCII stands for the American Standard Code Interchange Information.

3. Convert these measurements to the units indicated (approximate value only):
   a. 80 GB = ? B
   b. 4 TB = ? B
   c. 140 000 000 B = ? MB
   d. 3 000 000 000 B = ? GB

4. a. List the first twenty binary numbers from 0.
   b. What is the next binary number after 10011110?
   c. What is the ASCII code for the ‘per cent’ symbol?
   d. What is the character that is represented by the ASCII code 00100001?
Activities

5 Copy and complete the following table:

<table>
<thead>
<tr>
<th>Binary</th>
<th>128</th>
<th>64</th>
<th>32</th>
<th>16</th>
<th>8</th>
<th>4</th>
<th>2</th>
<th>1</th>
<th>Decimal number</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>00001011</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>b</td>
<td>01000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>01010111</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>11001001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>01100001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>00011111</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>11100001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>01111101</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>01000010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>j</td>
<td>01111111</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6 Copy and complete the following table:

<table>
<thead>
<tr>
<th>Decimal number</th>
<th>128</th>
<th>64</th>
<th>32</th>
<th>16</th>
<th>8</th>
<th>4</th>
<th>2</th>
<th>1</th>
<th>Binary code</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>65</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>01000001</td>
</tr>
<tr>
<td>b</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>101</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>119</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>j</td>
<td>167</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Development

7 Write a sentence for a friend in your class. Convert this sentence into ASCII code. Swap and decode your sentences.

8 Perform the following calculations.
   a Add 00101001₂ and 10101100₂.
   b Add 10010101₂ and 00010100₂.

9 Hexadecimal is an arithmetic system used by computers. Do research to find information on hexadecimal. Write a brief report that summarises the results of your investigation.
3.3 Data sources and data types

In this section we examine different sources of data and describe some of the different data types.

Primary sources

*Primary source* is when data is collected first hand. Many organisations prefer to generate their own primary data because they consider it meaningful and reliable, although it is time consuming and costly.

Surveys and interviews are a common method of acquiring primary data. These involve collecting sample characteristics, attitudes, behaviours and opinions. A survey is delivered by mail, face-to-face or online. An interview is a meeting of people face-to-face. Specific questions are asked to determine the interviewee’s opinions and feelings (see Figure 3.2).

A predesigned form or table is often used to collect primary data. A form is a document with blank spaces for information to be collected. For example, an online registration form is used to collect primary data for mailing lists. Forms and tables for data collection need to be carefully designed. It is often necessary to try to anticipate answers that people may give to individual questions or possible results for an event.

Secondary sources

*Secondary source* is the use of data collected or created by someone else. Secondary data is available as printed text and on the Internet.

- Printed text has been the main medium for spreading knowledge since the beginning of the seventeenth century. It can take many different forms, such as books, magazines, newspapers, newsletters, reports and journals. The sheer quantity of printed information means that an individual can absorb only a very small proportion. Searching for particular information from printed text involves using tables of contents, summaries and
indexes. Printed material may be accurate at the time of printing but it can quickly become out of date as society changes.

• The Internet has become a major source of data. It allows someone on the computer to access data stored on another computer. Websites present information on a particular topic (see Figure 3.3). Each single document is called a web page. In addition to the World Wide Web (www), data can be collected from newsgroups. A newsgroup is a discussion group on a specific topic. People access newsgroups each day and they provide a large diversity of opinion and knowledge. The Internet also allows data to be downloaded. It is possible to download files containing text, graphics, animation, video, sound and software.

Data types

Data exists in several different forms. The data type is the sort of data such as text and hypertext, graphics, audio, video and animation.

Text is data in the form of letters, numbers and other special characters. A word processor allows text to be entered and documents to be created (see Figure 3.4). It is used to write letters, reports, assignments, articles and books. Text is also entered into most software applications. Text is edited in many different ways such as deleting, inserting, moving and copying. The meaning of the text is determined by reading the text. Text is stored as an ASCII code. Some common file extensions for text include DOC, TXT and RTF.

The data type in some applications, such as a database, needs to be detailed. It requires the user to specify the exact nature of the data:

• alphanumeric—all keyboard characters
• numeric—numbers only
• alphabetic—letters only
• Boolean—true or false
• currency—money
• date—day, month, year.
Numeric data is often used to perform calculations in a database and a spreadsheet.

*Hypertext* is a system that allows documents to be cross-linked in such a way that the user can move from one document to another by clicking on text that is linked to that document. Hypertext links are usually identified using an underlined character and/or a different text colour. It is a common method of navigation used on the Internet. File extensions include HTM or HTML.

*Graphics* are pictures such as drawings, paintings or photographs. They can be stored, edited and transferred in similar ways to text. Graphics are created using graphics programs or captured using scanners, digital cameras or video cameras. All graphics are made up of tiny dots called pixels. A pixel (picture element) is the smallest part of the screen that can be controlled by the computer. Graphics are edited using graphic software. A graphic can be positioned anywhere on a page and can be cropped, resized or distorted.

There are two types of graphics: bit-mapped graphics and vector graphics. Editing objects in a bit-mapped graphic at an object level is often a problem. When you enlarge a bit-mapped graphic, you also enlarge each pixel within the image and you may create a staircase pattern called *aliasing*. Vector graphics are easily edited using a drawing program. The meaning of a graphic is determined by looking at the...
Some common file extensions for graphics include BMP, JPG, GIF, and TIF.

**Audio** is sound that has been digitised or represented in the form of digits or numbers. It may be a noise to get the participant’s attention or a voice to explain the operation of the software. Sounds are edited in many different ways, such as deleting sounds, changing the speed, adding an echo, overlaying (mixing) sound files and altering the quality of the sound file. The meaning of audio data is determined by listening to and interpreting the sounds (see Figure 3.6). Some common file extensions for sound include MP3, WAV, and MIDI.

**Video** is made up of frames that contain individual images. When the video is played, the frames are displayed in sequence. Video editing software is used to edit a video (see Figure 3.7). It involves obtaining the required video clip and adding text, audio, and graphics. The meaning of a video is determined by watching and listening over a period of time. Some common file extensions for video include MPG, AVI, and MOV.

**Animation** is the movement of an object. It is the result of a series of graphics or frames presented in rapid succession. Sometimes animations can present information more effectively than text or graphics. For example, an animation that shows the movement of blood through the body or the eruption of a volcano would be more effective than a graphic. The meaning of an animation is determined by watching it over a period of time. Some common file extensions for animation include SWF, MOV, and animated GIF.
Exercise 3.3

1. State the data type of the following files:
   a. photo.bmp
   b. music.mp3
   c. film.mpg
   d. data.doc
   e. cartoon.gif.

2. Copy and complete the following by replacing the letter in brackets with a suitable term:
   The (a) has become a major source of data. Websites present (b) on a particular topic. Each single document is called a (c). In addition to the World Wide Web, data is collected from (d).

3. True or false?
   a. Primary source is when data is collected or created by someone else.
   b. There are two types of graphics: BMP and JPG.
   c. A word processor allows text to be entered and documents to be created.
   d. Video is made up of frames that contain individual images.

4. a. Describe some of the different products that use printed text.
   b. What is hypertext?
   c. How are graphics created?
   d. How are sounds edited?
   e. What is animation?

Development

5. Choose a recent event that has been widely reported in newspapers and on television, radio and the Internet. Compare and contrast three different data sources on the event in terms of accuracy, currency and organisation.

6. The amount of information available in printed form and on the Internet is increasing at a very fast rate. Will the amount of information continue to increase in the future? Why? Many people are suffering from information overload. Outline some strategies that people could adopt to reduce information overload.

3.4 Primary storage

Data storage involves receiving and retaining data over a period of time. The amount of data storage is very important in determining the capabilities of a computer. Data storage is classified as primary storage or secondary storage.

Primary storage stores data and programs that need to be instantly accessible to the central processing unit (CPU). It is internal storage as it uses integrated circuits (silicon chips) located on the
motherboard. Primary storage is also known as main memory, primary memory, main storage or, simply, memory. It consists of RAM, cache and ROM.

**RAM**

RAM (random access memory) is where data and instructions are held temporarily. It depends on a supply of electricity to maintain data storage. When the power to the computer is shut off, everything in RAM is lost. In other words, RAM is *volatile* memory. The data in RAM can be accessed randomly. That is, a byte of memory is accessed without touching any of the other bytes. There are two main types of RAM:

- dynamic RAM (DRAM, pronounced ‘dee-ram’) is memory that must be constantly refreshed
- static RAM (SRAM, pronounced ‘ess-ram’) is memory that does not have to be updated or refreshed.

RAM is regarded as the working memory of the computer. Programs and files currently in use are stored in RAM. RAM is one of the factors that affects the computer’s performance. For this reason RAM manufacturers are continually inventing new designs to provide the fastest possible access times at the lowest possible cost. This has resulted in different types of DRAM and SRAM chips. Memory is upgraded by inserting DRAM chips into special slots on the motherboard (see Figure 3.8).

**Cache**

Cache (pronounced ‘cash’) is a temporary storage area used to store data that the computer can access quickly. There are two common types of cache:

- **Cache memory** is high-speed memory located between the CPU and RAM. It improves performance by using SRAM and reduces the need for the CPU to access the slower DRAM chip.
- **Disk cache** works the same way as cache memory. It stores the most recent data from the disk in RAM and when the CPU needs access to data from the disk, it first checks the disk cache. Disk caching improves the performance because accessing the data from RAM is a thousand times faster than accessing a disk.
ROM

ROM (read only memory) holds data and instructions that are fixed at the time of production and cannot be changed by the computer. It is permanent memory that only allows data to be retrieved (read) and not entered into storage (write). The software that is stored in ROM is called firmware. Storage of software within the ROM protects it from being damaged or changed. Most personal computers contain a small amount of ROM that stores important programs such as booting the computer and checking for input and output devices. The instructions that allow the computer to communicate with input and output devices are called the ROM BIOS (Basic Input/Output System). The computer manufacturer sets the actual contents of ROM and it is non-volatile. Non-volatile memory does not lose its contents when the power to the computer is turned off.

Exercise 3.4

1. What am I?
   a. The receiving and retaining of data over a period of time.
   b. A temporary storage area used to store frequently requested data and instructions.
   c. The software that is stored in ROM.
   d. Data and instructions that are fixed at the time of production and cannot be changed by the computer.

2. Copy and complete the following by replacing the letter in brackets with a suitable term:
   (a) storage is internal storage as it uses (b) circuits located on the motherboard. It consists of RAM, (c) and ROM. RAM is an abbreviation for (d) access memory.

3. Explain the difference between:
   a. dynamic RAM and static RAM
   b. RAM and ROM
   c. cache memory and disk cache
   d. volatile and non-volatile memory.

4. a. List some other terms for primary storage.
   b. Why is RAM described as volatile memory?
   c. How is memory upgraded?
   d. What is ROM BIOS memory?

Development

5. How much RAM is on your computer? Count the number of characters on a page of text (including spaces) and estimate the number of bytes required for this page. How many pages of text can be stored in your RAM? (Assume all the RAM is available to store the text.)
60

RAM is only a temporary storage and depends on a supply of electricity to maintain data storage. Design a poster to remind users about the importance of frequently saving their work.

7 The amount of primary storage is a determining factor in the speed of your computer. However, primary storage is more expensive than secondary storage. Some people argue that primary storage is more important than secondary storage. Do you agree with this statement? Give reasons for your answer.

3.5 Secondary storage

Secondary storage involves the use of a storage device not located on the computer’s motherboard. A storage device is any device that can store data and then allow it to be retrieved when required. It involves devices such as magnetic tape, magnetic disks, optical disks and flash memory.

Magnetic tape

Magnetic tape is a very long, thin strip of plastic, coated with a thin layer of magnetic material. Modern tape is wound on two reels inside a cartridge. Data is read from and written to tape using a tape drive that winds the tape from one reel to the other reel causing it to pass a read–write head. A magnetic tape can store large quantities of data in a small space at a relatively low cost. Magnetic tape is often used as a backup or copy of data (see Figure 3.9).

The main disadvantage with magnetic tape is that it uses sequential access to retrieve data. This form of access starts at the beginning of the tape and reads all of the data until the required item is found. Sequential access to data is slow, making magnetic tapes unsuitable for data that is often updated. Magnetic tapes have a variety of sizes and formats such as QIC tapes, DAT cartridges and 8 mm cartridges.

Figure 3.9 Magnetic tape is used for storage
**Magnetic disks**

A *magnetic disk* is a circular piece of metal or plastic, the surface of which has been coated with a thin layer of magnetic material. Magnetic disks use *random access* (or direct access) to retrieve data. This form of access allows data to be found directly without accessing all the previous data. Random access allows data to be retrieved much faster than sequential access used on magnetic tapes. A *disk drive* is a device on which a magnetic disk is mounted. The disk drive spins the disk and uses one or more heads to read and write data. If the disk is permanently attached to the drive it is called a fixed disk (hard disk).

A *floppy disk* (diskette) is a magnetic disk made of flexible plastic and covered with magnetic material. Even though floppies do not have a large storage capacity (1.4 MB) and are slower to access data than a hard drive, they are portable and cheap. The most common size of floppy disk is the 3.5 inch (9 cm). It is fully enclosed in a rigid plastic casing and is used for storing small files. To be used, a floppy disk must be inserted into the disk drive that is usually built into the system unit.

A *hard disk* is a magnetic disk made of metal or glass and covered with magnetic material (see Figure 3.10). It is rigid and not flexible like a floppy disk. This rigid construction allows it to be rotated ten to 100 times faster than a floppy disk, giving it faster access to data. Hard disks store more data than floppy disks because the data is stored more densely. A hard disk has a storage capacity measured in GB with larger capacities coming on the market regularly.

*Removable cartridges* are hard disks encased in a metal or plastic cartridge that are removed like a floppy disk. Removable cartridges are fast, though usually not as fast as fixed hard disks. They combine the best aspects of hard and floppy disks. There are two common types of removable cartridges used with personal computers called zip disks and jaz disks:
• **Zip disks** (or cartridges) are slightly larger than the 3.5-inch floppy disk and about twice as thick (see Figure 3.11). They can store 100 MB or 250 MB of data. Zip disks are popular for backing up hard disks and transporting files because they are relatively inexpensive.

• **Jaz disks** (or cartridges) are an upgrade of the zip disks. They store up to 2 GB of data. Jaz disks are suited to store large files for multimedia, graphics and sound.

---

**Optical media**

An optical disk is a storage medium in which data is read and written using laser technology. Optical disks have the capacity to store large amounts of data and are very reliable.

A **compact disc** (CD) is a polycarbonate disk, the surface of which is coated with a reflective layer of metal. The data is written to the CD by a high-powered laser that burns millions of tiny holes on the disk's surface, called pits. The standard CD is 12 centimetres wide and can store up to 800 MB. CDs use random access to retrieve data. There are many kinds of CDs such as CD-ROM, CD-R, and CD-RW:

• **CD-ROM** (compact disc—read only memory) contains data that cannot be changed. A CD-ROM drive is needed to read data. The larger the drive speed, the faster it can transfer data. CD-ROMs are convenient for storing data that remains constant.

• **CD-R** (compact disc—recordable) allows data to be recorded (or burned) but users cannot remove the information. CD-Rs require a CD-R drive to burn the data. They have become a popular storage medium. A CD-R drive allows users to create CD-ROMs and audio CDs.

• **CD-RW** (compact disc—rewritable) allows the user to write, erase and rewrite data. Erasing the disk is achieved by heating the surface and quickly cooling it. CD-RW is slower than a hard disk and after frequent use areas of the disk may become inaccessible.

---

**Figure 3.11** A zip disk is a large capacity removable storage device.
**Data handling**

**DVDs** (digital versatile disks) are optical disks that can store large amounts of data. DVDs are the same size as the standard CD but can store in excess of 17 GB. There are many kinds of DVDs such as DVD-ROM, DVD-RAM, DVD-R and DVD-RW. Each of these DVDs works in a similar way to CDs (discussed above) except that they can store more data. DVD-ROM drives can play DVD movies, read DVD data disks, read computer CDs, and play audio CDs.

**Flash memory**

Flash memory is an erasable memory chip used in portable computers, mobile phones, digital cameras and a range of other devices (see Figure 3.12). It retains data when the power is removed. The device is similar to a memory chip except it can be electrically erased. Flash memory is erased and written in fixed blocks ranging from 512 bytes to 256 KB. Flash chips are replacing the ROM BIOS chips so that the BIOS can be updated. It is often called flash BIOS. A flash memory card looks similar to a credit card and comes in a variety of formats, such as PC card, CompactFlash and SmartMedia.

Another type of flash memory device is a memory stick. A memory stick is a portable storage device that is connected using the USB port. It is compact, light, reliable and easy to handle. A memory stick allows easy transfer of image, audio and other data between computers and peripheral devices. Memory sticks are available in a range of different sizes including 1 GB.

**Exercise 3.5**

1. Copy and complete the following sentences:
   a. Magnetic tape is often used as a _____ or copy of the data.
   b. A _____ is a circular piece of metal or plastic, the surface of which has been coated with a thin layer of magnetic material.
   c. A floppy disk does not have a large storage capacity and is _____ to access data than a hard drive.
   d. Zip disks are slightly larger than a _____ and about twice as thick.
2 What am I?
   a Magnetic disk made of metal or glass and covered with magnetic material.
   b A type of removable cartridge that is suited to store large files.
   c A compact disc that allows the user to write, erase and rewrite data.
   d A hard disk encased in a metal or plastic cartridge that is removed like a floppy disk.

3 Unjumble these words:
   a fplypo idks
   b izp kdsi
   c afhls emyomr
   d sdki rdiev

4 a What is magnetic tape?
   b Explain the difference between sequential access and random access.
   c List three different types of magnetic disks.
   d Outline the advantages of using a memory stick for secondary storage.
   e Why are optical disks used for secondary storage?
   f What is a DVD?
   g Describe the appearance of flash memory.

Development
5 Formatting prepares a disk to accept data by organising it into sectors and tracks. It destroys any data that is currently on the disk. Is it appropriate to delete a file by formatting a disk? Give a reason. When is it appropriate to reformat a disk?

6 Do research into the experience offered by some museums and art galleries using the Internet and optical disks. Is it appropriate for artwork to be reproduced using information technology? Justify your argument by using some examples.

3.6 Data transmission and data compression

In this section we examine two methods for data transmission and explain some of the techniques used to compress data.

Data transmission
Data is transferred between computers and devices in two ways: serial transmission or parallel transmission. Serial transmission is the transmission of data one bit at a time through a single line. It is used to transmit data to peripheral devices, such as a keyboard and
mouse. Serial transmission is used on networks using either twisted-pair, coaxial or fibre optic cables (see Figure 3.13). Serial transmission has the advantage of being cheaper than parallel transmission as it uses only a single line. However, it has the disadvantage of being slower than parallel transmission. Serial transmission can be either synchronous or asynchronous.

- **Synchronous transmission** requires all the data to be sent at the same rate. The same number of bits is sent each second. This is synchronised by each device using a clock. Synchronous transmission is faster and more efficient than asynchronous transmission as there are no extra bits. It is used on larger computer systems.

- **Asynchronous transmission** is the sending of data by identifying each byte with special start and stop bits, and it has become the standard for personal computers. It uses an extra bit to check for errors in transmission.

Parallel transmission involves sending more than one bit at a time (usually 8 bits) simultaneously using separate lines (see Figure 3.14). It is much quicker than serial transmission because it can send more than one bit at a time. Parallel transmission is used only for distances less than a few metres, otherwise errors occur in transmitting the data and the cabling is too expensive. It is used inside the system unit and for connecting peripheral devices such as disk drives.

### Data compression

Data compression reduces the number of bits required to represent the information. It allows the user to store more data and makes data transmission faster. Compressed data must be decompressed to extract the original data. The amount a file is compressed is measured by the **compression ratio**. The compression ratio compares how much larger the uncompressed file is to the compressed file.
If a compressed file is half the size of the uncompressed file, the compression ratio is 2 to 1 (2:1). That is, two units of data in the uncompressed file equals one unit of data in the compressed file. A higher compression ratio indicates greater compression of data. For example, a compression ratio of 3:1 means the file is compressed smaller than a compression ratio of 2:1. A file compression ratio of 1:1 indicates the file has not been compressed.

**Lossy compression** permanently removes a number of bits from the file. The resulting file is smaller in size but not always identical to the original. However, audio and video files can be compressed using high compression ratios, with no noticeable change to the human ear or eye. JPEG and MPEG are common lossy compression formats for graphic and video files. An MPEG file can provide a compression ratio up to 200:1. This is only a fraction of the size of an uncompressed video file.

**Lossless compression** allows the original file to be recovered without loss of information. It works by replacing repeated data with something that takes up less room. For example, in this book the word ‘and’ appears many times. If you replace each ‘and’ with a one character symbol or token you save two-thirds of the space. Lossless compression is used mainly on text files such as word processing, spreadsheets and database files. Each file of a given type can have a different compression ratio depending on the contents of the file. A text file usually has a compression ratio of 3:1. Popular lossless compression systems include ZIP/PKZIP (Windows) and StuffIt (Macintosh).

**Exercise 3.6**

1. True or false?
   a. Serial transmission is much quicker than parallel transmission.
   b. Parallel transmission is used inside the system unit for connecting disk drives.
   c. Asynchronous transmission is faster and more efficient than synchronous transmission.
   d. Lossy compression is used to compress audio and video files.
   e. Lossless compression results in a file that is not always identical to the original file.
2 Copy and complete the following by replacing the letter in brackets with a suitable term:
   Data (a) reduces the number of bits required to represent information. The amount a file is compressed is measured by the (b). A compression ratio of (c) results in a file half the size of the original file. The higher the compression ratio the (d) the compression of data.

3 Unjumble these words:
   a allpreal
   b roatmnsisin
   c coonresmpi

4 a Explain the difference between serial and parallel transmission.
   b Describe synchronous transmission.
   c What is lossy compression?
   d How does lossless compression work?
   e List some of the popular lossless compression systems.

5 A method for detecting errors in the asynchronous transmission is to attach an additional bit to the binary code. This additional bit is called a parity bit. Do research into parity checking. How does it work? Give an example.

6 ‘Data compression is important for download times.’ Explain this statement. Examples should be used to illustrate your answer.

7 Construct a table to compare text, graphic, audio and video data types. For each data type consider the typical file size, file format and software program to edit the data.

3.7 Data security

Data security is a series of safeguards to protect data. It has become a major issue for governments and organisations. The cost of replacing data that is deliberately or accidentally damaged or lost can be enormous. The war against terrorism has highlighted the need for data security.

Need for data security

Data security is needed to prevent the following threats:
• information theft—stealing data from one organisation and selling it to another organisation
• financial theft—illegal transfer of money from one account to another
• information alteration—changing or deleting existing data
• unlawful access—unauthorised access to data in a computer system
• burglary—stealing a computer with its data
• vandalism—intentional damage to a computer system
• viruses—creating or introducing computer viruses
• software piracy—illegal copying of software
• natural disaster—events such as a fire or flood that could destroy data
• accidental—unintentional loss or damage to data.

All of these threats, except accidental or natural disaster, involve a person committing a computer crime. A computer crime is any illegal activity that involves the use of computer technology. People who gain illegal access to data in computer systems are called hackers. Hackers and computer crime are a growing problem that present our society with some difficult issues.

Even though hundreds of cases of computer crimes are reported each year, many crimes go unreported. There are several reasons for this: organisations are reluctant to face adverse publicity, publicising the event could give others ideas, and in some cases the crime is not discovered. Are organisations encouraging computer crime by not reporting it? Are organisations that do not report computer crime guilty of a crime themselves?

A second issue is that computer crime is not held in the same light by the community as other criminal activities such as armed robbery. Whereas the community holds an armed robber in fear or contempt, a hacker is often regarded as clever and the crime is not regarded as dangerous or threatening. If a burglar and a hacker stole $300 000 from a business, does it make any difference how the money was taken? Should the penalties for both crimes be the same?

Computer criminals have stolen people’s personal data and effectively become that person. They have obtained credit card numbers, driver’s licence numbers, phone numbers and even birth details. People often provide this data by completing online forms. However, computer criminals have obtained sensitive information
by searching through people’s rubbish. Unfortunately an increase in e-commerce has resulted in an increase in online fraud. People need to be cautious when providing personal details.

**Basic security methods**

Data security is a series of safeguards to protect data. Some of the basic security methods include the following:

- **Passwords** are secret words or numbers that are typed on the keyboard to gain access to the system. It is important to choose a password that is not obvious. Good data protection systems change passwords often, so that only authorised persons can have access to certain data.

- **Personal objects** carried to gain access to the information system such as a key, plastic card, or badge. They are often used together with a personal identification number (PIN).

- **Biometric devices** are used to verify personal characteristics such as fingerprints, hand size, signature, eye and voice.

- **Data encryption** is used to prevent data from being intercepted during transmission. Encryption is the process of coding data and decryption the process of changing it back. It is the most effective way to achieve data security during the transmission of data. Data is coded, transmitted, and then converted back to its original form.

- **Firewalls** are used on the Internet or any network to check all incoming data for the purpose of verification and authentication (see Figure 3.16). A firewall aims to protect a computer system from hackers trying to access sensitive information. On large systems more than one firewall is necessary because barriers need to be placed at all critical points.

- **Waste** is secured, since discarded printouts are sources of information to unauthorised persons. This kind of waste can be secured by using a shredder.

*Figure 3.16 A firewall is used to protect data*
The weakest link in the security of any computer system is the people in it. Employees need to be carefully screened. This is difficult, as well-respected and generally honest employees can commit computer crimes.

Backup procedures need to be secure and reliable. A regular plan of copying and storing data will guard against data loss. Organisations should keep backup copies in a fireproof safe or offsite.

A certificate or digital signature is an encrypted code that identifies a particular person or website. It aims to ensure that unauthorised persons cannot access data.

---

**Case study**

**Project: DataProtect**

DataProtect is a business that specialises in the security of data. It protects organisations from viruses, disk crashes and system bombs. An employee of DataProtect has been asked to write a brief document on simple precautions to avoid data loss on a PC. The problem was solved using the four stages in project development.

- **Define and analyse the problem:** The employee discussed the size and format of the document with the manager. Each employee was surveyed on possible content for the document. A project plan was written and submitted for approval to the manager.

- **Design possible solutions:** Three draft solutions to the problem were completed. Each solution contained an overview for data security but emphasised a different aspect of data security. Each solution was discussed with the manager who preferred the solution that emphasised backup procedures.

- **Produce the solution:** The document was created using Microsoft Word. It contained simple precautions on protecting passwords, regular backups, use of anti-virus software, sharing data, installing a firewall, downloading data and opening unknown emails. A draft copy of the document was printed and checked. The final copy was presented to the manager.

- **Evaluate the solution:** The manager was pleased with the document. It will be part of the documentation given to customers of DataProtect. Customers will be asked for feedback on the document as part of the evaluation.

**Tasks**

1. Do research into current precautions for data security on a PC. Write a brief report that summarises your investigation.
2. Create a document that would be appropriate for this project.
Exercise 3.7

1 Copy and complete the following sentences:
   a _______ is the stealing of data from one organisation and selling it to another organisation.
   b People who gain illegal access to data in a computer system are called _______.
   c _______ are used to verify personal characteristics such as fingerprints.
   d The opposite process to encryption is called _______.

2 True or false?
   a The cost of replacing data that is deliberately or accidentally damaged or lost can be enormous.
   b Information alteration is the illegal transfer of money from one account to another.
   c The weakest link in the security of any computer system is the people in it.

3 What am I?
   a The illegal copying of software.
   b An illegal activity that involves the use of computer technology.
   c Security method that prevents data from being intercepted during transmission.
   d Secret words or numbers that are typed on the keyboard to gain access to the system.

4 a What is data security?
   b Why are computer crimes often not reported?
   c What is the purpose of a firewall?
   d Do you think a hacker and an armed robber who stole the same amount of money should receive the same penalties for their crimes? Explain your answer.

Development

5 The Internet is basically insecure. It was originally designed for communication between people who were trusted. Any unprotected computer system accessing the Internet is completely vulnerable. A firewall is an essential tool for all Internet users.’ Investigate the protection offered by three firewall programs. How would you rate the protection offered by these programs? Write a report of your investigation.

6 A survey on Internet and network security found more than 36 per cent of local IT departments had experienced severe breaches. Do research into computer crime. Describe some examples of security breaches. What penalties have been imposed?

7 Websites keep a record of your visits and may be able to identify you by name. ‘Anonymisers’ hide your identity and also encrypt the addresses you visit so your ISP cannot keep a record of them. Research the Internet for more information on ‘anonymisers’. Summarise your investigation.
Part A: Multiple choice questions
Select the alternative (a), (b), (c) or (d) that best answers each question.

1 What is data that is represented in the form of digits or numbers called?
   a Information
   b Binary
   c Analog
   d Digital

2 Which of the following binary numbers represents the decimal number 97?
   a 01100001
   b 01100010
   c 00110001
   d 10100010

3 Which of the following is not a data type?
   a Text
   b Photograph
   c Video
   d Audio

4 How many bytes is a gigabyte?
   a Trillion bytes
   b Billion bytes
   c Thousand bytes
   d Million bytes

5 What is primary storage called where data and instructions are held temporarily?
   a RAM
   b Hard disk
   c ROM
   d Motherboard

6 Which of the following is not a secondary storage device?
   a Disk cache
   b Floppy disk
   c Removable cartridge
   d Memory stick

7 What is an optical disc called that allows the user to write, erase and rewrite data?
   a CD-A
   b CR-ROM
   c CD-R
   d CD-RW

8 What type of transmission involves sending more than one bit at a time?
   a Parallel
   b Serial
   c Asynchronous
   d Synchronous

9 What is data compression called that allows the original file to be recovered?
   a WINZIP
   b Recapture
   c Lossy
   d Lossless

10 Which of the following is not a basic security method?
    a Password
    b Firewall
    c Virus
    d Backup

Part B: Matching the term
For each of the following statements (1 to 10), select from the list of terms (a to j) the one that most closely fits the statement.

Statements
1 Data that is represented by using continuous variable physical quantities.
   2 A standard code that changes letters, numbers and symbols into an 8-bit code.
Part C: Extended response questions
Write at least one paragraph for each of the following:

1 'People can collect huge amounts of data but not satisfy their information needs.' Explain this statement.

2 How does a personal computer represent the letter ‘Z’? Your explanation should refer to binary code and ASCII.

3 Construct a table that lists, defines and gives an example of the five data types.

4 Explain the difference between primary and secondary storage. What is the meaning of the acronyms RAM and ROM?

5 Why is magnetic tape still used for secondary storage? How is data stored on this medium?

6 Data transmission between computers and devices is completed in two ways. Describe the differences in the two methods of data transmission.

7 Why has data security become an important issue? List six basic security methods.
Project: Data handling

Investigate the latest information technology used in data handling. Work in teams and write a report to summarise one recent development in data handling. The report should contain relevant graphics and follow good design principles. Suggested topics include the latest RAM or ROM chips, modern devices used for secondary storage, popular data compression techniques or recent data security measures.