

**22**

330en22

# FUNDAMENTALS OF DATA STRUCTURE

In the previous module you have learnt about C++ programming. Now you will learn about the fundamentals of data structure. Each organisation, regardless of its size or purpose, generates data to keep a record of events and transactions that take place within the business. Generating and organising this data in a useful way is called data processing. In this lesson, you will learn about various terms such as data, information, data processing and data processing system.



## OBJECTIVES

After reading this lesson, you will be able to:

- define the concepts of data, information and data processing;
- explain various data processing activities;
- utilise data processing cycle;
- explain data elements, records, files and databases.

## 22.1 DATA

The word “data” is the plural of datum, which means fact, observation, assumption or occurrence. More precisely, data are representations of facts pertaining to people, things, ideas and events. Data are represented by symbols such as letters of the alphabets, numerals or other special symbols.

## 22.2 DATA PROCESSING

Data processing is the act of handling or manipulating data in some fashion. Regardless of the activities involved in it, processing tries to assign meaning to data. Thus, the ultimate goal of processing is to transform data into information.



Notes

Data processing is the process through which facts and figures are collected, assigned meaning, communicated to others and retained for future use. Hence we can define data processing as a series of actions or operations that converts data into useful information. We use the term ‘data processing system’ to include the resources that are used to accomplish the processing of data.

**22.3 INFORMATION**

Information, thus can be defined as “data that has been transformed into a meaningful and useful form for specific purposes”. In some cases data may not require any processing before constituting information. However, generally, data is not useful unless it is subjected to a process through which it is manipulated and organised, its **content is analyzed** and **evaluated**. Only then data becomes information.

There is no hard and fast rule for determining when data becomes information. A set of letters and numbers may be meaningful to one person, but may have no meaning to another. Information is identified and defined by its users.

For example, when you purchase something in a departmental store, a number of data items are put together, such as your name, address articles you bought, the number of items purchased, the price, the tax and the amount you paid. Separately, these are all data items but if you put these items together, they represent information about a business transaction.

**22.4 DATA PROCESSING ACTIVITIES**

As discussed above, data processing consists of those activities which are necessary to transform data into information. Computer is not the only tool used for data processing; it can be done without computers also. However, computers have outperformed people for certain tasks. There are some other tasks for which computer are a poor substitute for human skill and intelligence.

Regardless to the type of equipment used, various functions and activities which need to be performed for data processing can be grouped under five basic categories as shown in Fig. 22.1.

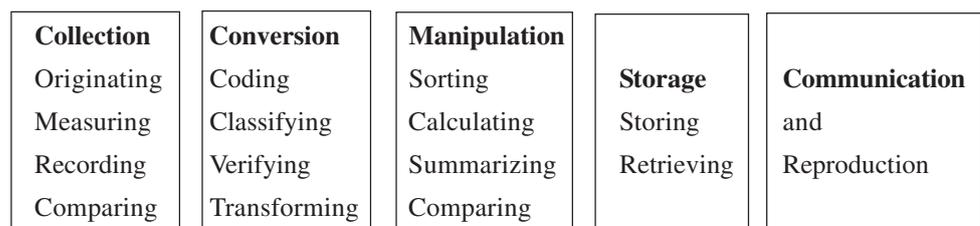


Fig. 22.1: Data Processing Activities

We will now describe each of these activities:

Deptt. 5 Employee Time Sheet for Week ending 21/07/2001							
Hours Worked							
Emp No.	Name	Mon	Tues	Wed	Thurs	Fri	Total Hours
110	Hariom	8	8	8	8	8	40
150	Dinesh	8	8	8	8	6	38
160	Rajesh	8	4	8	8	8	36
170	Pankaj	8	6	8	5	8	35

Fig. 22.2: The source document for a payroll application is the weekly time sheet

### (a) Collection

Data originates in the form of events transaction or some observations. This data is then recorded in some usable form. Data may be initially recorded on paper source documents 22.2 and then converted into a machine usable form for processing. Alternatively, it may be recorded by a direct input device in a paperless, machine-readable form. Data collection is also termed as **data capture**.

### (b) Conversion

Once the data is collected, it is converted from its source documents to a form that is more suitable for processing. The data is first codified by assigning identification codes. A code comprises of numbers, letters, special characters, or a combination of these. For example, an employee may be allotted a code as 52-53-162, his category as A class, etc. It is useful to codify data, when data requires classification. To classify means to categorize, i.e., data with similar characteristics are placed in similar categories or groups. For example, one may like to arrange accounts data according to account number or date. Hence a balance sheet can easily be prepared.

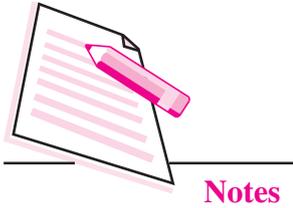
After classification of data, it is verified or checked to ensure the accuracy before processing starts. After verification, the data is transcribed from one data medium to another. For example, in case data processing is done using a computer, the data may be transformed from source documents to machine sensible form using magnetic tape or a disk.

### (c) Manipulation

Once data is collected and converted, it is ready for the manipulation function which converts data into information. Manipulation consists of the activities like sorting, calculating, summarizing and comparing.



Notes



Notes

**Sorting**

It involves the arrangement of data items in a desired sequence. Usually, it is easier to work with data if it is arranged in a logical sequence. Most often, the data are arranged in alphabetical sequence. Sometimes sorting itself will transform data into information. For example, a simple act of sorting the names in alphabetical order gives meaning to a telephone directory. The directory will be practically worthless without sorting. Business data processing extensively utilises sorting technique. Virtually all the records in business files are maintained in some logical sequence. Numeric sorting is common in computer-based processing systems because it is usually faster than alphabetical sorting.

**Calculating**

Arithmetic manipulation of data is called calculating. Items of recorded data can be added to one another, subtracted, divided or multiplied to create new data as shown in Fig. 22.2(a). Calculation is an integral part of data processing. For example, in calculating an employee’s pay, the hours worked multiplied by the hourly wage rate gives the gross pay. Based on total earning, income-tax deductions are computed and subtracted from gross-pay to arrive at net pay.

Weekly Payroll Summary Report 07/07/2005					
Dept.	Employee Number	Name	Hours Worked	Pay Rate	Gross Wages
2	170	Pankaj	34	10.00	340.00
	175	Rekha	32	9.00	288.00
	158	Param	20	5.00	100.00
	160	Rajesh	36	8.00	288.00
	165	Sanju	45	9.00	405.00
	159	Savini	25	3.00	75.00
Department 2 Total			192		1496.00
4	110	Hari Om	33	7.75	255.75
	115	Vimla	40	9.25	370.00
Department 4 Total			73		625.75
8	150	Dinesh	38	7.00	266.00
	155	Upma	40	9.45	390.00
	156	Arunima	32	8.50	272.00
	157	Ritwik	40	8.00	320.00
Department 8 Total			150		1,248.00
Company Total			415		3,369.75

Fig. 22.2(a) : Weekly Payroll Summary



Notes

### Summarizing

To summarize is to condense or reduce masses of data to a more usable and concise form as shown in Fig. 22.2(b). For example, you may summarize a lecture attended in a class by writing small notes in one or two pages. When the **data involved** is numbers, you summarize by counting or accumulating the total of the data in a classification or by selecting strategic data from the mass of data being processed. For example, the summarizing activity may provide a general manager with sales-totals by major product line, the sales manager with sales totals by individual salesman as well as by the product line and a salesman with sales data by customer as well as by product line.

Department Payroll Summary for Week of 07-07-2005		
Department	Hours	Gross Wages
2	192	1,496.00
4	73	625.75
8	150	1,248.00
Company Total	415	3,369.75

Fig. 22.2(b) : Department Payroll Summary

### Comparing

To compare data is to perform an evaluation in relation to some known measure. For example, business managers compare data to discover how well their companies are doing. They compare current sales figures with those for last year to analyze the performance of the company in the current month.

#### (d) Managing the Output Results

Once data has been captured and manipulated following activities may be carried out :

##### Storing

To store is to hold data for continued or later use. Storage is essential for any organised method of processing and re-using data. The storage mechanisms for data processing systems are file cabinets in a manual system, and electronic devices such as magnetic disks/magnetic tapes in case of computer based system. The storing activity involves storing data and information in organised manner in order to facilitate the retrieval activity. Of course, data should be stored only if the value of having them in future exceeds the storage cost.



Notes

### Retrieving

To retrieve means to recover or find again the stored data or information. Retrieval techniques use data storage devices. Thus data, whether in file cabinets or in computers can be recalled for further processing. Retrieval and comparison of old data gives meaning to current information.

### (e) Communication

Communication is the process of sharing information. Unless the information is made available to the users who need it, it is worthless. Thus, communication involves the transfer of data and information produced by the data processing system to the prospective users of such information or to another data processing system. As a result, reports and documents are prepared and delivered to the users. In electronic data processing, results are communicated through display units or terminals.

### Reproduction

To reproduce is to copy or duplicate data or information. This reproduction activity may be done by hand or by machine.



### INTEXT QUESTIONS 22.1

1. Differentiate between the following :
  - (a) Data and Information
  - (b) Data processing and Data processing system
2. Define the following terms briefly:
  - (a) Sorting
  - (b) Summarizing

### 22.5 THE DATA PROCESSING CYCLE

The data processing activities described above are common to all data processing systems from manual to electronic systems. These activities can be grouped in four functional categories, viz., data input, data processing, data output and storage, constituting what is known as a data processing cycle.

#### (i) Input

The term input refers to the activities required to record data and to make it available for processing. The input can also include the steps necessary to check, verify and validate data contents.

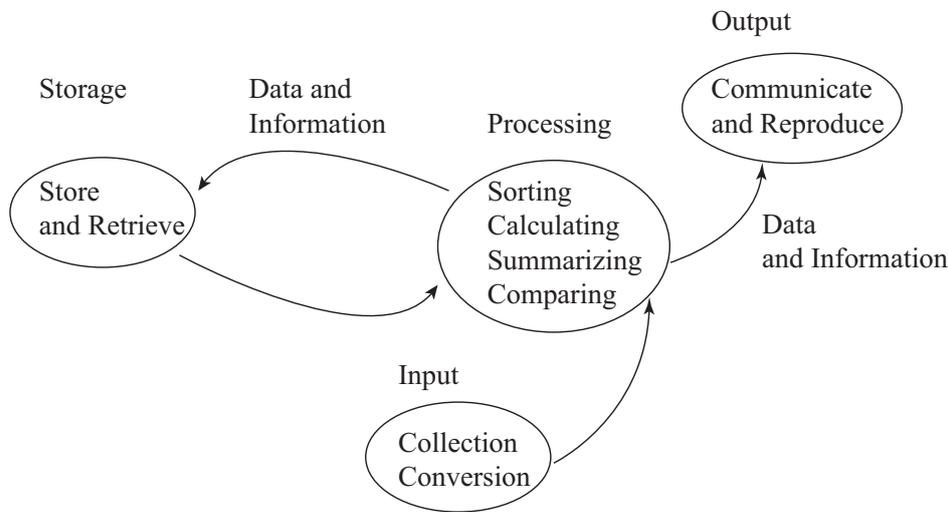


Fig. 22.3: Data Processing Cycle

**(ii) Processing**

The term processing denotes the actual data manipulation techniques such as classifying, sorting, calculating, summarizing, comparing, etc., that convert data into information.

**(iii) Output**

It is a communication function which transmits the information, generated after processing of data, to persons who need the information. Sometimes output also includes decoding activity which converts the electronically generated information into human-readable form.

**(iv) Storage**

It involves the filing of data and information for future use. The above mentioned four basic functions are performed in a logical sequence as shown in Fig. 22.3 in all data processing systems.

## 22.6 COMPUTER PROCESSING OPERATIONS

A computer can perform only the following four operations which enable computers to carry out the various data processing activities we have just discussed.

**(a) Input/Output operations**

A computer can accept data (input) from and supply processed data (output) to a wide range of input/output devices. These devices such as keyboards, display screens, and printers make human-machine communication possible.



Notes



Notes

**(b) Calculation and text manipulation Operations**

Computer circuits perform calculations on numbers. They are also capable of manipulating numerics and other symbols used in text with equal efficiency.

**(c) Logic/Comparison Operations**

A computer also possesses the ability to perform logic operations. For example, if we compare two items represented by the symbols A and B, there are only three possible outcomes.

A is less than B ( $A < B$ ); A is equal to B ( $A = B$ ); or A is greater than B ( $A > B$ ). A computer can perform such comparisons and depending on the result, follow a predetermined path to complete its work. This ability to compare is an important property of computers.

**(d) Storage and Retrieval Operations**

Both data and program instructions are stored internally in a computer. Once they are stored in the internal memory, they can be called up quickly or retrieved, for further use.

**22.7 DATA PROCESSING SYSTEM**

The activity of data processing can be viewed as a “system”. According to James O’Brien a system can be defined as “a group of interrelated components that seeks the attainment of a common goal by accepting inputs and producing outputs in an organised process”. For example, a production system accepts raw material as input and produces finished goods as output.

Similarly, a data processing system can be viewed as a system that uses data as input and processes this data to produce information as output.



There are many kinds of data processing systems. A manual data processing system is one that utilizes tools like pens, and filing cabinets. A mechanical data processing system uses devices such as typewriters, calculating machines and book-keeping machines. Finally, electronic data processing uses computers to automatically process data.

## 22.8 DATA ORGANISATION

Having discussed the Data Processing Cycle (also called Information Processing Cycle) and the components of a computer, we will now describe how data is organised before processing on a computer. Data can be arranged in a variety of ways, but a hierarchical approach to organisation is generally recommended.

### Data Item

A data item is the smallest unit of information stored in computer file. It is a single element used to represent a fact such as an employee's name, item price, etc. In a payroll application, the employee number 170 is a data item. PANKAJ, the name is a data item.

### Field

Data items are physically arranged as fields in a computer file. Their length may be fixed or variable. Since all individuals have 3 digit employee numbers, a 3-digit field is required to store the particular data. Hence, it is a fixed field. In contrast, since customer's name vary considerably from one customer to another, a variable amount of space must be available to store this element. This can be called as variable field.

### Record

A record is a collection of related data items or fields. Each record normally corresponds to a specific unit of information. For example, various fields in the record, illustrated in Fig. 22.4 are employee number, employee's name, basic salary and house rent allowance. This is the data used to produce the payroll register report. The first record contains all the data concerning the employee PANKAJ. The second record contains all the data concerning the employee REKHA. Each subsequent record contains all the data for a given employee. It can be seen how each related item is grouped together to form a record.

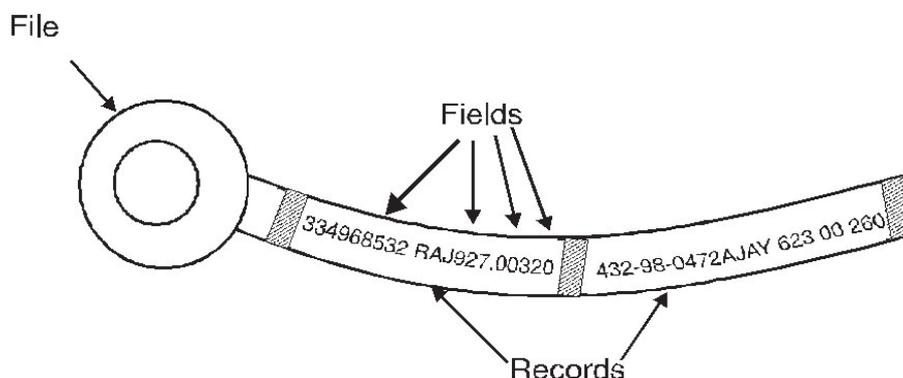


Fig. 22.4: Data Record



Notes



Notes

**File**

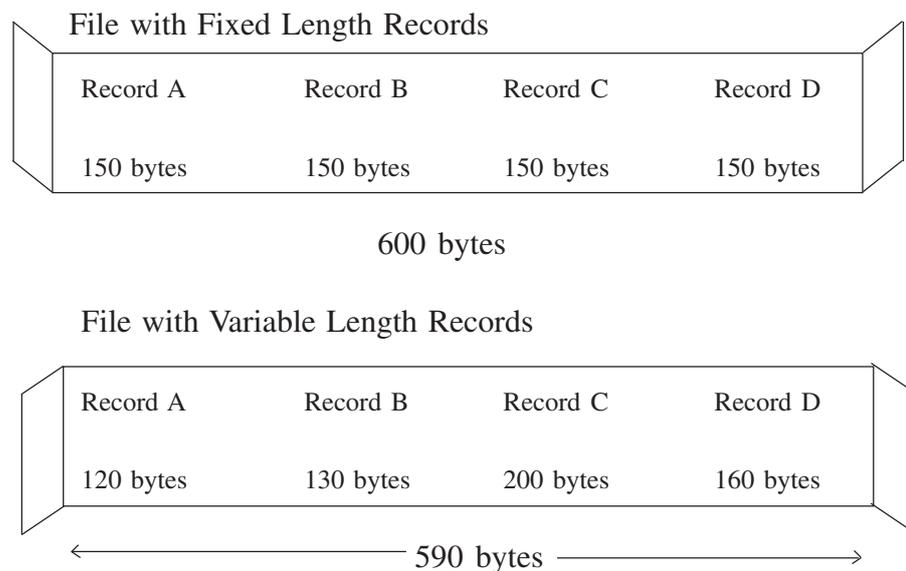
The collection of records is called a file. A file contains all the related records for an application. Therefore, the payroll file shown in Fig. 22.5 contains all records required to produce the payroll register report. Files are stored on some medium, such as magnetic disk, flash drive or CD ROM.

**Database**

The collection of related files is called as database. A database contains all the related files for a particular application.

**22.9 VARIABLE AND FIXED LENGTH RECORDS**

Records can be of fixed or variable length as depicted in Fig. 22.5.



**Fig. 22.5: Fixed and variable length records.**

**Fixed Length Records**

In this case, all the records in a file have the same number of bytes. Such a file is called a **flat file**. If all the records are expected to contain essentially the same quantity of data, then fixed length records are used.

**Variable Length Records**

In this case, records vary in length. Use of variable length records conserves storage space when the quantity of information, of various records in a file, differs significantly.

**22.10 LOGICAL VERSUS PHYSICAL RECORD**

A logical record contains all the data related to a single entity. It may be a payroll record for an employee or a record of marks secured by a student in a particular examination. A physical record refers to a record whose data fields are stored physically next to one another. It is also the amount of data that is treated as a single unit by the input-output device. Portions of the same logical record may be located in different physical records or several logical records may be located in one physical record. For example, in case of magnetic tape, number of logical records are stored in the form of a block to increase the data transfer speed and this block is referred to as a physical record as shown in Fig. 22.6

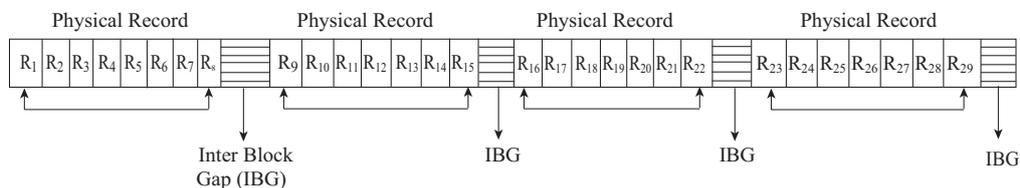


Fig. 22.6: Physical and logical Record

**INTEXT QUESTIONS 22.2**

1. Identify various data processing activities.
2. Define the various steps of data processing cycles.
3. State whether the following statements are true or false:
  - (i) All the records in a file have the same number of bytes, are called as fixed length.
  - (ii) A physical record refers to a record whose data fields are stored physically next to one another.
  - (iii) Data items are physically arranged as fields in a computer file. Their length is fixed.

**WHAT YOU HAVE LEARNT**

- Data is a fact, observation, assumption or occurrence.
- Data processing is a series of actions or operations that converts data into useful information.



Notes

## MODULE – 4

Database Concepts,  
Web Designing

### Fundamentals of Data Structure



Notes

- Data that has been transformed into a meaningful and useful form for specific purposes is called as information.
- Data processing activities can be grouped as collection, conversion, manipulation, communication and reproduction.
- Data input, data processing, data output and storage constitutes data processing cycle.
- Data item is the smallest unit of information stored in computer file.
- A record is a collection of related data items or fields.
- The collection of records is called as file.
- The collection of related files is called as database.



### TERMINAL EXERCISE

1. Define the terms data, data processing and information.
2. Identify various activities involved in manipulation.
3. Write down the various computer processing operations briefly.
4. How data is organised before processing on a computer starts? Discuss briefly.



### ANSWERS TO INTEXT QUESTIONS

#### 22.1

1. (a) The word “data” is the plural of datum which means fact, observation, assumption or occurrence. On the other hand, information can be defined as data that has been transformed into a meaningful and useful form for specific purposes.  
(b) Data processing is the process through which facts and figures are collected, assigned meaning, communicated further and retained for future use. It is a series of actions or operations that converts data into useful information. In data processing system, we include the resources that used to accomplish the processing of data.

2. (a) It involves the arrangement of data items in a desired sequence.
- (b) It means to condense or reduce masses of data to a more usable and concise form.

### 22.2

1. Data processing activities are grouped under following five basic categories.
  - (i) Collection
  - (ii) Conversion
  - (iii) Manipulation
  - (iv) Storage and retrieval
  - (v) Communication
2. The various steps involved in data processing cycle are as follows:
  - (i) Data input
  - (ii) Data processing
  - (iii) Output
  - (iv) Storage
3. (i) True
- (ii) True
- (iii) False



Notes