

# CHAPTER ONE:

## INTRODUCTION TO DATA STRUCTURES:

- **Data Structure**: This is a systematic way to organize data in order to use it efficiently.
- **Interface**: Each data structure has an interface. Interface represents the set of operations that a data structure supports. An interface only provides the list of supported operations, type of parameters they can accept and return type of these operations.
- **Implementation** – Implementation provides the internal representation of a data structure. Implementation also provides the definition of the algorithms used in the operations of the data structure.

### Characteristics of a Data Structure

- **Correctness** – Data structure implementation should implement its interface correctly.
- **Time Complexity** – Running time or the execution time of operations of data structure must be as small as possible.
- **Space Complexity** – Memory usage of a data structure operation should be as little as possible.

### Need for Data Structure

As applications are getting complex and data rich, there are three common problems that applications face now-a-days.

- **Data Search** – Consider an inventory of 1 million- $(10^6)$  items of a store. If the application is to search an item, it has to search an item in 1 million- $(10^6)$  items every time slowing down the search. As data grows, search will become slower.
- **Processor speed** – Processor speed although being very high, falls limited if the data grows to billion records.
- **Multiple requests** – As thousands of users can search data simultaneously on a web server, even the fast server fails while searching the data.

Data can be organized in a data structure in such a way that all items may not be required to be searched, and the required data can be searched almost instantly.

### Execution Time Cases

There are three cases which are usually used to compare various data structure's execution time in a relative manner.

- **Worst Case** – This is the scenario where a particular data structure operation takes maximum time it can take. If an operation's worst case time is  $f(n)$  then this operation will not take more than  $f(n)$  time where  $f(n)$  represents function of  $n$ .
- **Average Case** – This is the scenario depicting the average execution time of an operation of a data structure. If an operation takes  $f(n)$  time in execution, then  $m$  operations will take  $mf(n)$  time.

- **Best Case** – This is the scenario depicting the least possible execution time of an operation of a data structure. If an operation takes  $f(n)$  time in execution, then the actual operation may take time as the random number which would be maximum as  $f(n)$ .

## Basic Terminology

- **Data** – Data are values or set of values.
- **Data Item** – Data item refers to single unit of values.
- **Group Items** – Data items that are divided into sub items are called as Group Items.
- **Elementary Items** – Data items that cannot be divided are called as Elementary Items.
- **Attribute and Entity** – An entity is that which contains certain attributes or properties, which may be assigned values.
- **Entity Set** – Entities of similar attributes form an entity set.
- **Field** – Field is a single elementary unit of information representing an attribute of an entity.
- **Record** – Record is a collection of field values of a given entity.
- **File** – File is a collection of records of the entities in a given entity set.

## LOCAL ENVIRONMENT SETUP

If you are still willing to set up your environment for C programming language, you need the following two tools available on your computer,

- (a) Text Editor and
- (b) The C Compiler.

### Text Editor

This will be used to type your program. Examples of few editors include Windows Notepad, OS Edit command, Brief, Epsilon, EMACS, and vim.

### The C Compiler

The source code written in the source file is the human readable source for your program. It needs to be "compiled", to turn into machine language so that your CPU can actually execute the program as per the given instructions.

This C programming language compiler will be used to compile your source code into a final executable program.

Most frequently used and free available compiler is GNU C/C++ compiler. Otherwise, you can have compilers either from HP or Solaris if you have respective Operating Systems (OS).

### Installation on UNIX/Linux

If you are using Linux or UNIX, then check whether GCC is installed on your system by entering the following command from the command line

### Installation on Mac OS

If you use Mac OS X, the easiest way to obtain GCC is to download the Xcode development environment from Apple's website and follow the simple installation instructions. Once you have Xcode setup, you will be able to use GNU compiler for C/C++.

## **Installation on Windows**

To install GCC on Windows, you need to install MinGW. To install MinGW, go to the MinGW homepage, [www.mingw.org](http://www.mingw.org) and follow the link to the MinGW download page. Download the latest version of the MinGW installation program, which should be named MinGW-<version>.exe.

## **BASIC CONCEPTS OF ALGORITHMS**

An algorithm is a sequence of steps that must be carried out in order to accomplish a particular task. Three things are to be considered while writing an algorithm: input, process, and output. The input that we give to an algorithm is processed with the help of the procedure and finally, the algorithm returns the output.

### **HISTORY OF ALGORITHMS.**

The origin of the word algorithm is indirectly linked to India. A scholar in the, ‘House of wisdom’ in Baghdad, Abu Abdullah Muhammad Ibn Musa Al Khwarizmi, wrote a book about Indian numerals in which rules of performing arithmetic with such numerals were discussed. These rules were referred to as ‘algorism’ from which the word algorithm was derived. His book was translated into Latin in the 18th century. This was followed by the invention of Boolean algebra by George Boole and the creation of language in special symbols by Frege. The concept of algorithms, given its present form by a genius named Alan Turing, helped in the inception of artificial intelligence. Algorithms have been used for long in mathematics and computer science. Euclid’s theorem and the algorithm of Archimedes to calculate the value of ‘Pi’ are classic examples of algorithms. These events reinforced the belief that if a task is to be performed, then it must be performed with a predefined sequence of steps that are unambiguous and efficient. However, this belief would be challenged in the late 20th century with the introduction of non-deterministic algorithms. Not only in mathematics and computer science are there algorithms, they are part of our daily lives. When a person is taught how to make tea, even then an algorithm is edified. Algorithms are camouflaged as directions and rules, which form the basis of our existence. The challenge, however, is to make these algorithms efficient and robust.

### **Features/characteristics of algorithms.**

Not all procedures can be called an algorithm. An algorithm should have the following characteristics –

- ❖ **Unambiguous** – Each of its steps (or phases), and their inputs/outputs should be clear and must lead to only one meaning.
- ❖ **Input** – An algorithm should have 0 or more well-defined inputs.
- ❖ **Output** – An algorithm should have 1 or more well-defined outputs, and should match the desired output.
- ❖ **Finiteness** – Algorithms must terminate after a finite number of steps.
- ❖ **Feasibility** – Should be feasible with the available resources.

- ❖ **Independent** – An algorithm should have step-by-step directions, which should be independent of any programming code.
- ❖ It should be efficient both in terms of memory and time.

From the data structure point of view, following are some important categories of algorithms.

- **Search** – Algorithm to search an item in a data structure.
- **Sort** – Algorithm to sort items in a certain order.
- **Insert** – Algorithm to insert item in a data structure.
- **Update** – Algorithm to update an existing item in a data structure.
- **Delete** – Algorithm to delete an existing item from a data structure.

## IMPORTANCE OF ALGORITHMS

- It helps in enhancing the thinking process. They are like brain stimulants that will give a boost to our thinking process.
- It helps in solving many problems in computer science, computational biology, and economics.
- Without the knowledge of algorithms we can become a coder but not a programmer.
- A good understanding of algorithms will help us to get a job. There is an immense demand of good programmers in the software industry who can analyse the problem well.
- Genetic algorithms and randomized approach will help us to retain that job in the changing market.

## WAYS OF WRITING AN ALGORITHM

There are three basic ways of writing algorithms in programming. They include:

1. English like algorithm
2. Flowcharts
3. pseudocodes

### English-Like Algorithm.

**Problem** – Design an algorithm to add two numbers and display the result.

```

Step 1 – START
Step 2 – declare three integers a, b & c
Step 3 – define values of a & b
Step 4 – add values of a & b
Step 5 – store output of step 4 to c
Step 6 – print c
Step 7 – STOP

```

### Alternatively:

```

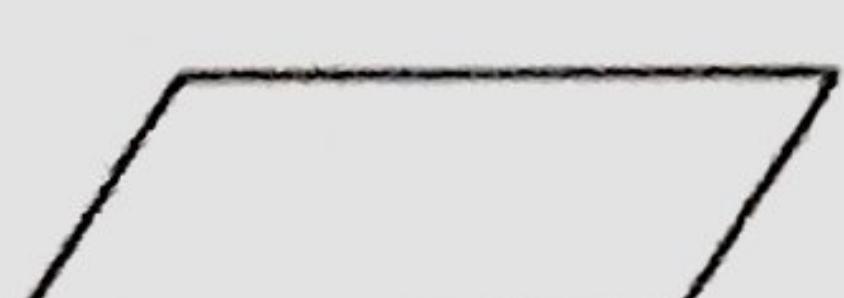
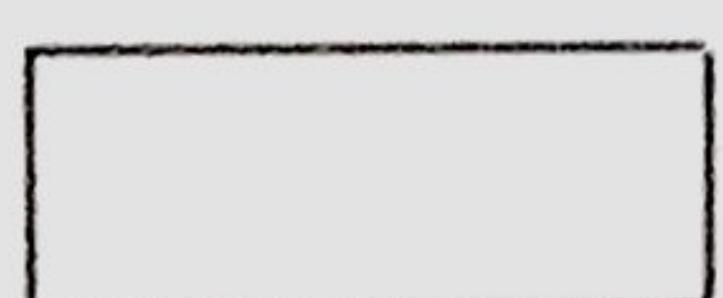
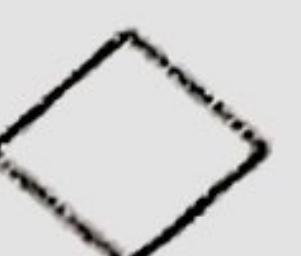
Step 1 – START ADD
Step 2 – get values of a & b
Step 3 – c ← a + b
Step 4 – display c
Step 5 – STOP

```

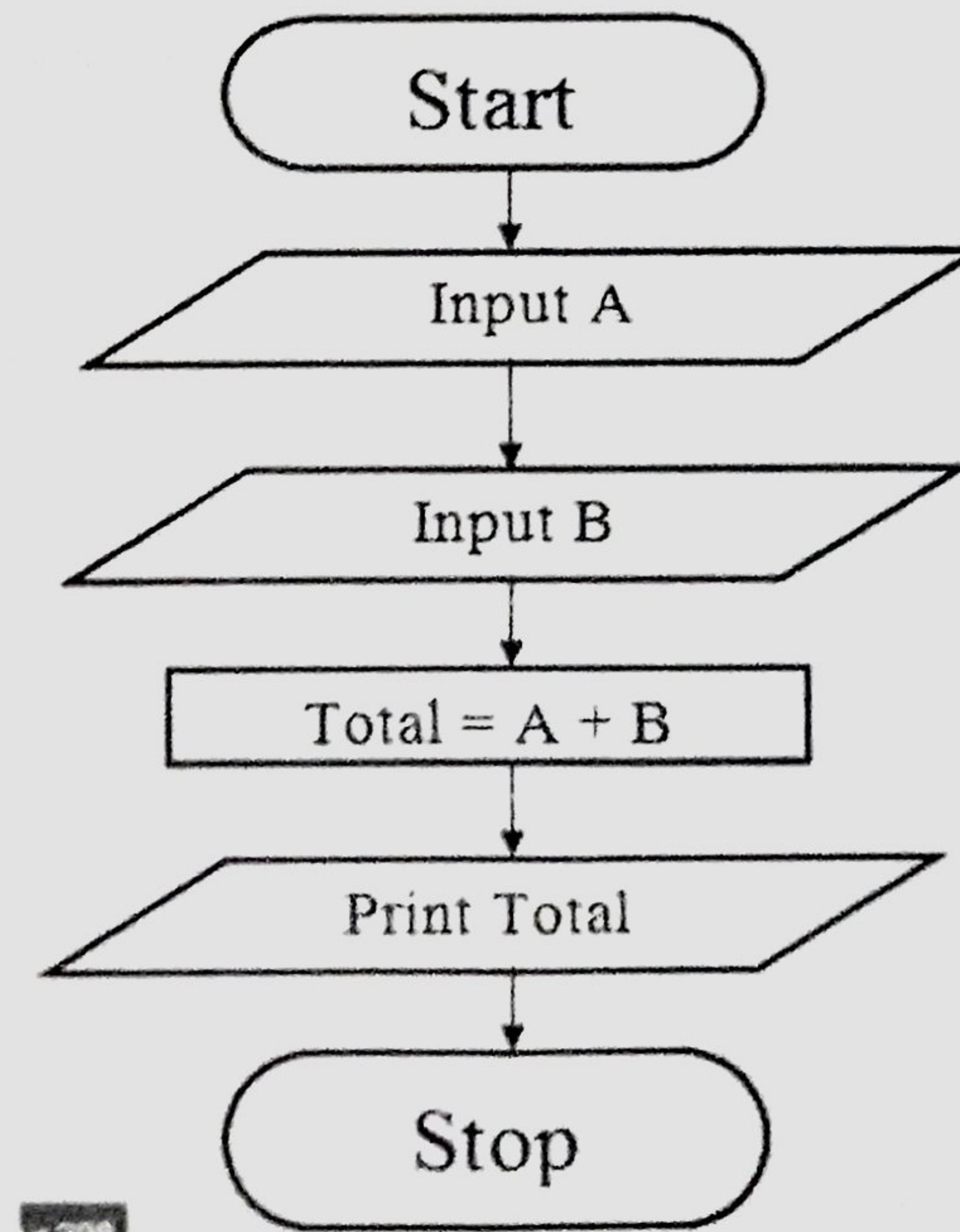
## Flowchart.

This is a graphical representation of a computer program in relation to its sequence of functions (as distinct from the data it processes). Flowcharts use simple geometric shapes to depict processes and arrows to show relationships and process/data flow.

Flowcharts pictorially depict a process.

S. No.	Name	Element Representation	Meaning
1.	Start/End		An oval is used to indicate the beginning and end of an algorithm.
2.	Arrows		An arrow indicates the direction of flow of the algorithm.
3.	Connectors		Circles with arrows connect the disconnected flowchart.
4.	Input/Output		A parallelogram indicates the input or output.
5.	Process		A rectangle indicates a computation.
6.	Decision		A diamond indicates a point where a decision is made.

**Problem – Design an algorithm to add two numbers and display the result.**



**Pseudocode.**

This is a notation resembling a simplified programming language, used in program design. The pseudocode has an advantage of being easily converted into any programming language. This way of writing algorithm is most acceptable and most widely used. In order to write a pseudocode, one must be familiar with the conventions of writing it.

### THEY INCLUDE:

1. Single line comments start with //
2. Multi-line comments occur between /\* and \*/
3. Blocks are represented using brackets. Blocks can be used to represent compound statements or the procedures.
4. Statements are delimited by semicolon.
5. Assignment statements indicates that the result of evaluation of the expression will be stored in the variable.
6. The boolean expression 'x > y' returns true if x is greater than y, else returns false.
7. The boolean expression 'x < y' returns true if x is less than y, else returns false.
8. The boolean expression 'x <= y' returns true if x is less than or equal to y, else returns false.
9. The boolean expression 'x >= y' returns true if x is greater than or equal to y, else returns false.
10. The boolean expression 'x != y' returns true if x is not equal to y, else returns false.
11. The boolean expression 'x == y' returns true if x is equal to y, else returns false.
12. The boolean expression 'x AND y' returns true if both conditions are true, else returns false.
13. The boolean expression 'x OR y' returns true if any of the conditions is true, else returns false.
14. The boolean expression 'NOT y' returns true if the result of x evaluates to false, else returns false.
15. if< condition >then< statement >
16. This condition is an enhancement of the above 'if' statement. It can also handle the case where the condition isn't satisfied.

**Problem –** Design an algorithm to add two numbers and display the result.

1. BEGIN
2. NUMBER a, b, sum
3. OUTPUT("input number a")
4. INPUT a
5. OUTPUT("input number b")
6. INPUT a
7. Sum = a + b
8. OUTPUT sum
9. END

### Data Definition

Data Definition defines a particular data with the following characteristics.

- **Atomic** – Definition should define a single concept.
- **Traceable** – Definition should be able to be mapped to some data element.

- **Accurate** – Definition should be unambiguous.
- **Clear and Concise** – Definition should be understandable.

## **Data Object**

Data Object represents an object having a data.

## **Data Type**

Data type is a way to classify various types of data such as integer, string, etc. which determines the values that can be used with the corresponding type of data, the type of operations that can be performed on the corresponding type of data. There are two data types –

- Built-in Data Type
- Derived Data Type

## Built-in Data Type

Those data types for which a language has built-in support are known as Built-in Data types. For example, most of the languages provide the following built-in data types.

- Integers
- Boolean (true, false)
- Floating (Decimal numbers)
- Character and Strings

## Derived Data Type

Those data types which are implementation independent as they can be implemented in one or the other way are known as derived data types. These data types are normally built by the combination of primary or built-in data types and associated operations on them. For example –

- List
- Array
- Stack
- Queue

## Basic Operations

The data in the data structures are processed by certain operations. The particular data structure chosen largely depends on the frequency of the operation that needs to be performed on the data structure.

- Traversing
- Searching
- Insertion
- Deletion
- Sorting
- Merging

## Array Data Structures

Array is a very basic data structure provided by every programming language. Let's talk about an example scenario where we need to store ten employees' data in our C/C++ program including name, age and salary. One of the solutions is to declare ten different variables to store employee name and ten more to store age and so on. Also you will need some sort of mechanism to get information about an employee, search employee records and sort them. To solve these types of problem C/C++ provide a mechanism called Arrays.

An array is simply a number of memory locations, each of which can store an item of data of the same data type and which are all referenced through the same variable name. Array may be defined abstractly as finite order set of homogeneous elements. So we can say that there are finite numbers of elements in an array and all the elements are of same data type. Also array elements are ordered i.e. we can access a specific array element by an index.

## Terms

- **Element** – Each item stored in an array is called an element.
- **Index** – Each location of an element in an array has a numerical index, which is used to identify the element.

The general form of declaring a simple (one dimensional)  
Array is array \_type variable name[array\_size];

in your C/C++ program you can declare an array like

```
int Age[10];
```

Here array type declares base type of array which is the type of each element in array. In our example array type is int and its name is Age. Size of the array is defined by array\_size i.e. 10. We can access array elements by index, and first item in array is at index 0. First element of array is called lower bound and its always 0. Highest element in array is called upper bound. In C programming language upper and lower bounds cannot be changed during the execution of the program, so array length can be set only when the program is written.

Array has 10 elements

Age 0	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7	Age 8	Age 9
30	32	54	32	26	29	23	43	34	5

Note: One good practice is to declare array length as a constant identifier. This will minimise the required work to change the array size during program development. Considering the array we declared above we can declare it like.

```
#define NUM_EMPLOYEE 10 int Age[NUM_EMPLOYEE];
```

### Initialising an array

Initialisation of array is very simple in C programming. There are two ways you can initialise arrays.

- Declare and initialise array in one statement.
- Declare and initialise array separately.

Look at the following C code which demonstrates the declaration and initialisation of an array.

```
int Age [5] = {30, 22, 33, 44, 25};
```

### Alternatively,

```
int Age [5];
Age [0]=30;
Age [1]=22;
Age [2]=33;
Age [3]=44;
Age [4]=25;
```

Array can also be initialised in a way that array size is omitted, in such case compiler automatically allocates memory to array.

```
int Age [ ] = {30, 22, 33, 44, 25};
```

Let's write a simple program that uses arrays to print out number of employees having salary more than 3000.

### Array in C Programming

```
#include <windows.h>
#include <stdio.h>
#include <stdlib.h>
#define NUM_EMPLOYEE 10
int main(int argc, char *argv[]){
int Salary[NUM_EMPLOYEE], lCount=0,gCount=0,i=0;
printf("Enter employee salary (Max 10)\n ");
for (i=0; i<NUM_EMPLOYEE; i++){
printf("\nEnter employee salary: %d - ",i+1);
scanf("%d",&Salary[i]); } for(i=0; i<NUM_EMPLOYEE; i++)
{
```

```

if(Salary[i]<3000)
lCount++;
else gCount++;
}
printf("\nThere are %d employee with salary more than 3000\n",gCount);
printf("There are %d employee with salary less than 3000\n",lCount); printf("Press ENTER to
continue...\n");
getchar();
return 0;
}

```

## Array in C++ Programming

```

#include <cstdlib>
#include <iostream>
#define NUM_EMPLOYEE 10 using namespace std;
int main(int argc, char *argv[])
{
    int Salary[NUM_EMPLOYEE], lCount=0,gCount=0,i=0;
    cout << "Enter employee salary (Max 10) " << endl;
    for (i=0; i<NUM_EMPLOYEE; i++)
    {
        cout << "Enter employee salary: - " << i+1 << endl;
        cin >> Salary[i];
    }
    for(i=0; i<NUM_EMPLOYEE; i++)
    {
        if(Salary[i]<3000) lCount++;
        else
            gCount++;
    }
    cout << "There are " << gCount << " employee with salary more than 3000" << endl
    << "There are " << lCount << " employee with salary less than 3000" << endl;
    system("PAUSE");
    return EXIT_SUCCESS;
}

```

## Basic Operations

Following are the basic operations supported by an array.

- **Traverse** – print all the array elements one by one.
- **Insertion** – Adds an element at the given index.
- **Deletion** – Deletes an element at the given index.
- **Search** – Searches an element using the given index or by the value.
- **Update** – Updates an element at the given index.

## How to declare and initialize multi-dimensional arrays

Often there is need to manipulate tabular data or matrices. For example if employee salary is increased by 20% and you are required to store both the salaries in your program. Then you will need to store this information into a two dimensional arrays. C/C++ gives you the ability to have arrays of any dimension.

```
int Salary[10][2];
```

This defines an array containing 10 elements of type int (integer). Each of these elements itself is an array of two integers. So to keep track of each element of this array we have to use two indices. One is to keep track of row and other is to keep track of column.

### Initialising multidimensional arrays

Multidimensional arrays can also be initialised in two ways just like one dimensional array. Two braces are used to surround the row element of arrays. If you are initialising more than one dimension then you will have to use as many braces as the dimensions of the array are.

```
int Salary [5][2] = {
```

```
{2300, 460},
```

```
{3400, 680},
```

```
{3200, 640},
```

```
{1200, 240},
```

```
{3450, 690};
```

```
int Salary [5][2] ={0}; //This will initialise all the array elements to 0
```

```
int Salary [5][2];
```

```
Salary [0][0]=2300;
```

```
Salary [1][0]=3400;
```

```
Salary [2][0]=3200;
```

```
Salary [3][0]=1200;
```

```
Salary [4][0]=3450;
```

```
Salary [0][1]=460;
```

```
Salary [1][1]=680;
```

```
Salary [2][1]=640;
```

```
Salary [3][1]=240;
```

```
Salary [4][1]=690;
```

The code below demonstrates two dimension arrays. It uses the same example of employee salary to increment it by 20% and adds it to actual salary then print current salary, increment and new salary. Two dimensional Array in C Programming

```
#include <windows.h>
#include <stdio.h>
#include <stdlib.h>
#define NUM_EMPLOYEE 10
int main(int argc, char *argv[]){ //initialise Salary of each employee
int Salary[NUM_EMPLOYEE][2]={
{2300,0},
{3400,0},
{3200,0},
{1200,0}, {3450,0},
{3800,0}, {3900,0},
{2680,0},
{3340,0},
{3000,0} };
int lCount=0,gCount=0,i=0;
for(i=0; i<NUM_EMPLOYEE; i++){
Salary[i][1] = ((Salary[i][0]*20)/100); }
printf("Initial Salary + Increment = Total Salary\n"); for (i=0; i<NUM_EMPLOYEE; i++)
{
printf("%d\t%d\t%d\n",Salary[i][0],Salary[i][1],Salary[i][0]+Salary[i][1]);
}
printf("Press ENTER to continue...\n"); getchar(); return 0; } Two dimensional array in C++
} printf("Press ENTER to continue...\n"); getchar(); return 0; } Two dimensional array in C++
#include <cstdlib> #include <iostream> #define NUM_EMPLOYEE 10 using
Programming namespace std; int main(int argc, char *argv[]){
//initialise Salary of each employee
int Salary[NUM_EMPLOYEE][2]={
{2300,0},
```

{3400,0},

{3200,0},

{1200,0},

{3450,0}

Using C++, write a program that implements algorithm for inserting data elements into one dimensional array? Solution:

```
#include <iostream.h>
```

```
include name space std
```

```
{
```

```
int i=0,x=0;
```

```
int a{};
```

```
for (i=0;i<10;i++)
```

```
cin>>a>>endl;
```

```
{
```

```
for(x=0;x<i;x++)
```

```
{
```

```
cout<<a<<endl;
```

```
}
```

## Review Questions

EXERCISE 1. Using C++, write a program that implements algorithm for inserting data elements into one dimensional array?

EXERCISE 2. Illustrate the concept of single linked list using C++ programming language.

EXERCISE 3. Write a C++ program that implements the algorithms for pushing, popping and deleting data elements from the stack data structure EXERCISE 4. Illustrate how queue data structure is different from stack data structure

EXERCISE 5. Construct a binary tree and apply the three traversal techniques on the following expression  $(A+B)*(C-D)$

EXERCISE 6. Discuss the concept of graph data structure?