
INFORMATION TECHNOLOGY

for

P.C.C.

&

I.P.C.C



PROF. ZULESH DEDHIA

FLOWCHARTS

In order to communicate with the computer, a computer language is used. However, since the computer does not have its own brains, it is foolish to expect common sense from it. Therefore, to get the computer to perform one task, a series of instructions are to be given in a computer language.

A program is a set to instructions given to the computer by the user, in a computer language, to get the computer to perform a certain task. These instructions are to be defined in a proper sequence and not haphazardly given. This specific sequence of instructions designed to get the desired result is called an Algorithm.

Diagrammatical presentation of an Algorithm is known as Flowchart. Thus, a Flowchart is a pictorial presentation of the sequence of steps involved in solving a problem.

SYMBOLS USED IN FLOWCHART:

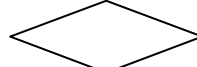
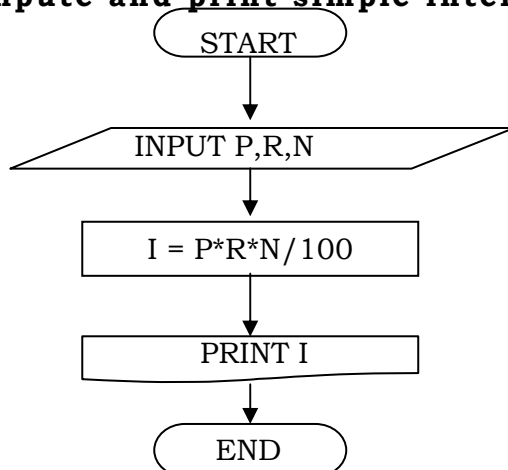
1. START/END	
2. INSTRUCTION FLOW LINE	
3. INPUT/OUTPUT	
4. PROCESS	
5. DECISION	
6. ON-PAGE CONNECTOR	
7. OFF-PAGE CONNECTOR	
8. PRINT	
9. DISPLAY	
10. ONLINE STORAGE/FLOPPY DISC	
11. MAGNETIC TAPE / SEQUENTIAL ACCESS STORAGE	

Illustration: Compute and print simple interest.



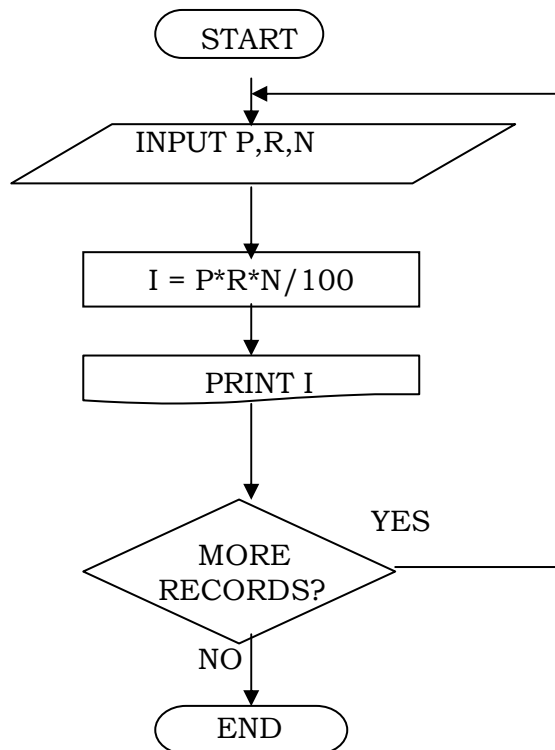
RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

P,R,N & I are variables as they can assume any value, i.e., their value is not fixed.

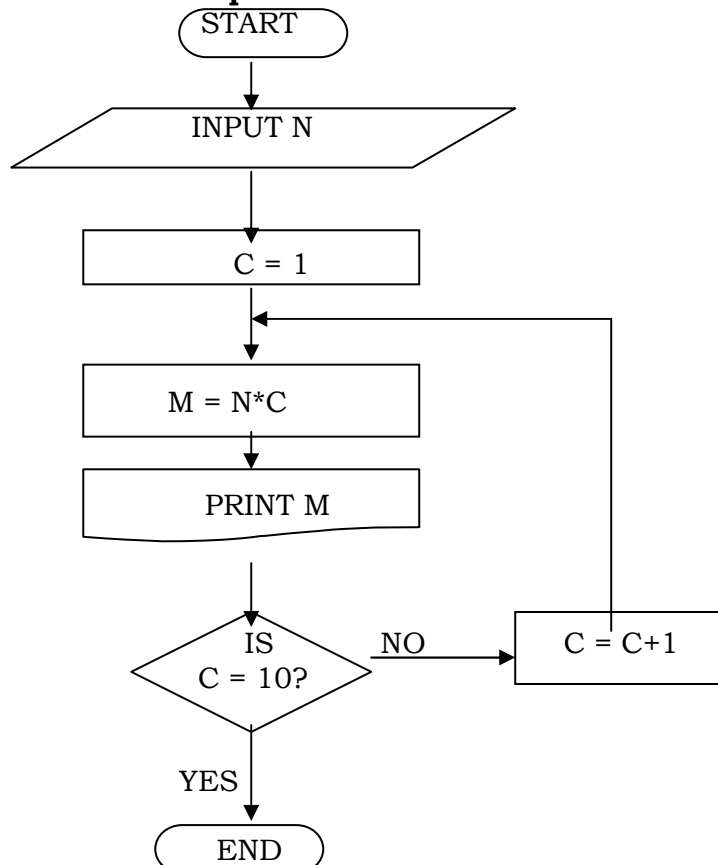
LOOP

When some portion of a program is to be executed repetitively, we create a loop. This loop may be set according to the happening of certain condition or it may be preset for specified number of times.

Loop set according to the happening of certain condition – An illustration:



Loop set for specific number of times - An illustration:
Find and print first 10 multiples of a number:

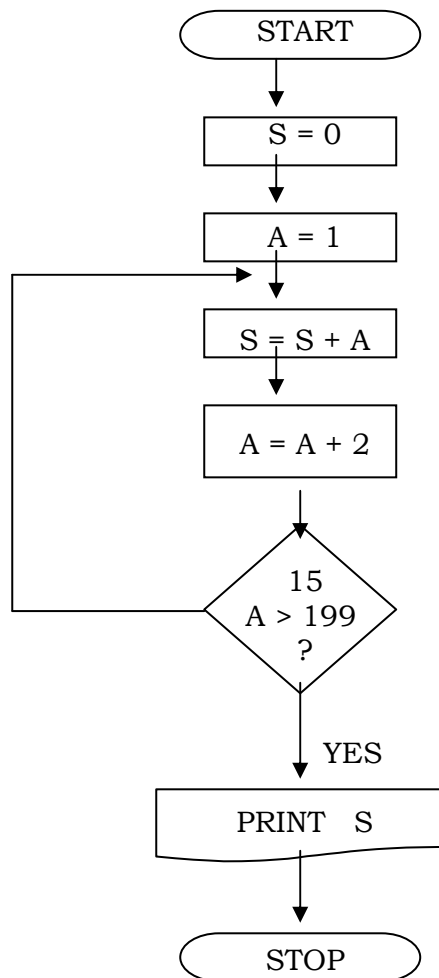


RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

COUNTER: A counter is a variable used to store the count of the loop. When the counter reaches the pre-set value, the program gets out of the loop. The use of the counter may be of two types- one for performing the loop a specified number of times and other as a variable for computation purpose. In the above example, a counter "C" has been used and it serves both the purposes.

ACCUMULATOR: A memory which is used to store intermediate results during the execution of loop is called accumulator. It is generally used for computing cumulative results of mathematical expression within the loop. Accumulators are usually given an initial value 0 or 1 depending upon the logic of the program. If the accumulator is used to accumulate a sum, initial value of accumulator is given 0. If the product of a series is to be computed, the initial value of accumulator is given as 1.

Draw a flowchart to find the sum of first 100 odd numbers.

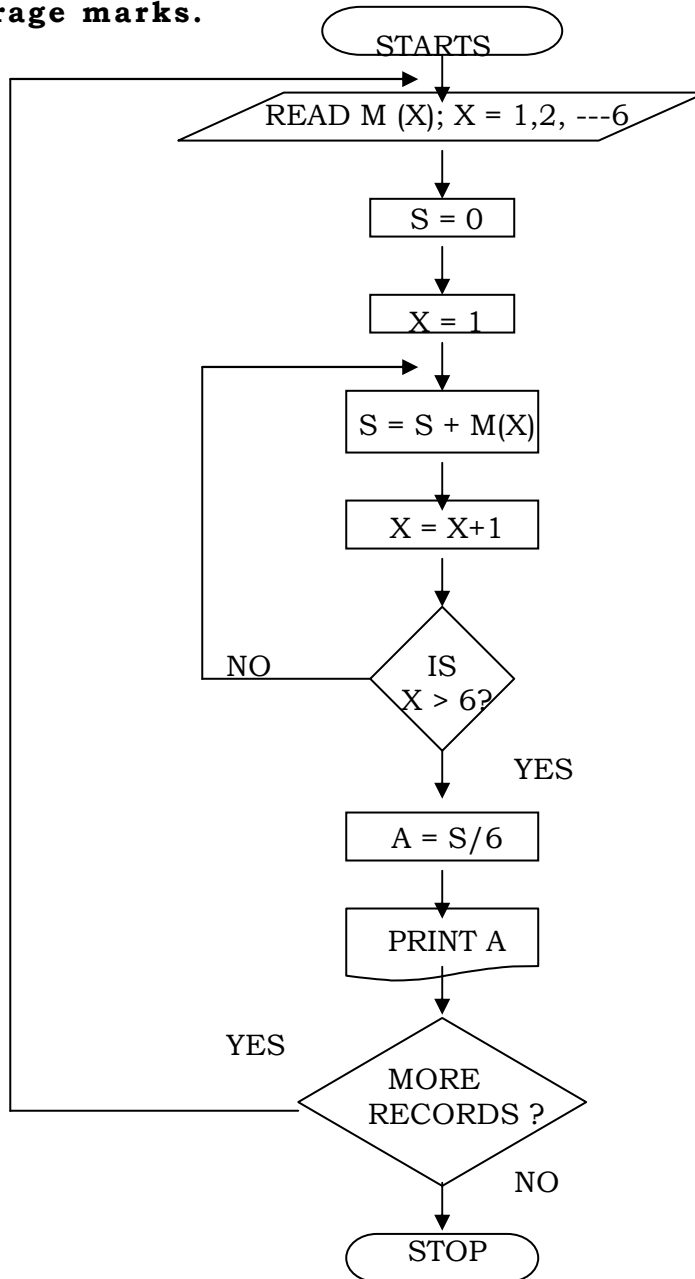


Here, S has been used as an accumulator and its initial value is taken as 0.

RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

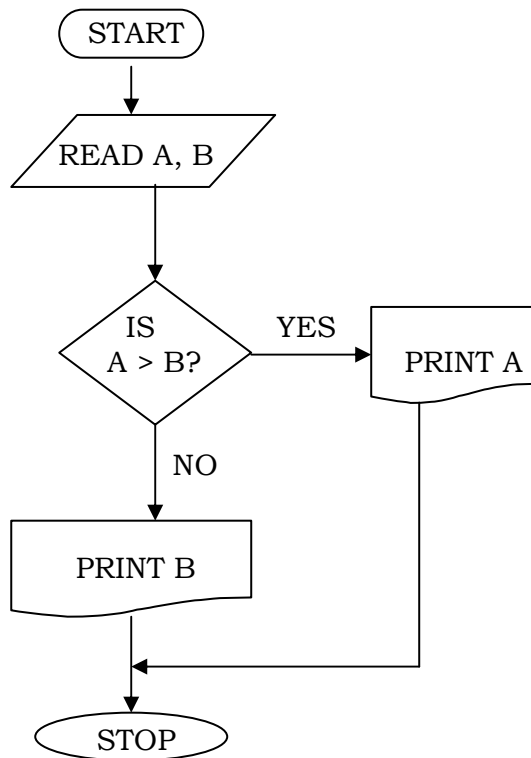
ARRAYS: A variable can store and represent only one value. Some times it is required to store a list of values and refer then by a common name. E.g. production of 12 months of a year is a set of similar items and we would like to use the same name to refer to this list of values. We may use P1, P2, P12 as variable names. This ensures that different variables belonging to same class are called by a single name. Example P. The 1,2,3,..... used in the above variables are called subscripts and P1, P2 etc. are subscripted variables.

Illustration: Draw a flowchart to accept marks of students in six subjects and print average marks.

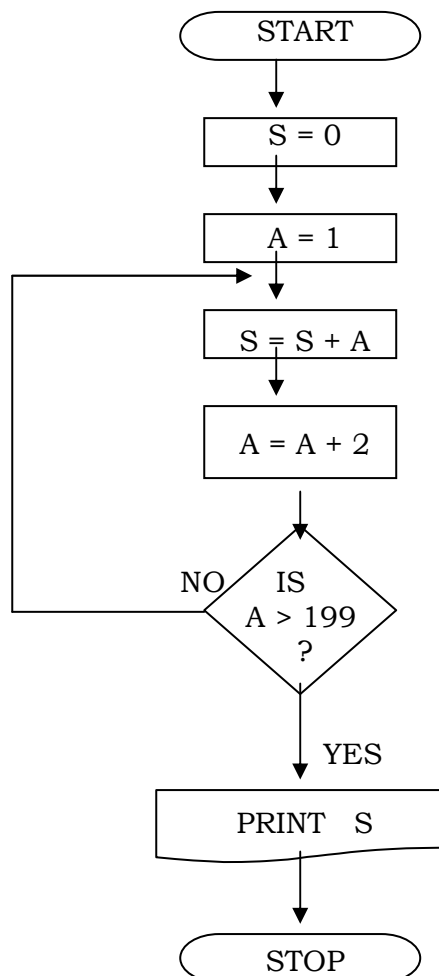


Solved Problems On Flowcharts:

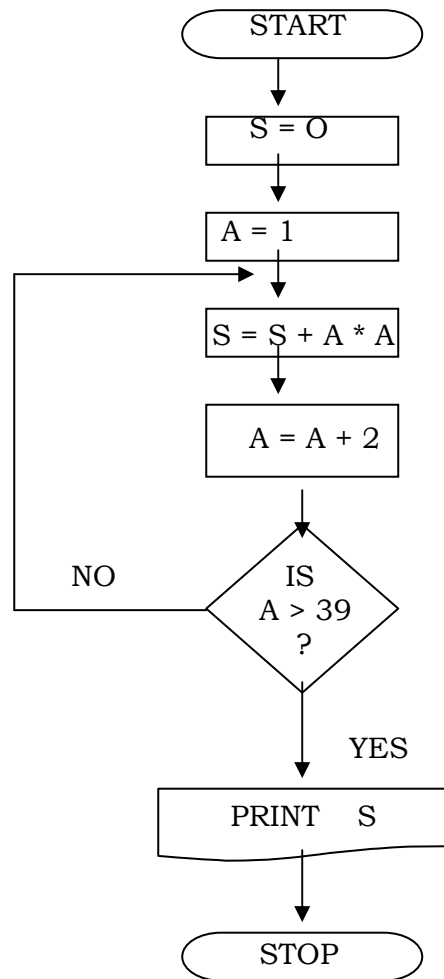
Q1. Draw a flowchart to print the higher of any given two unique numbers.



Q2. Draw a flowchart to find the sum of first 100 odd numbers.

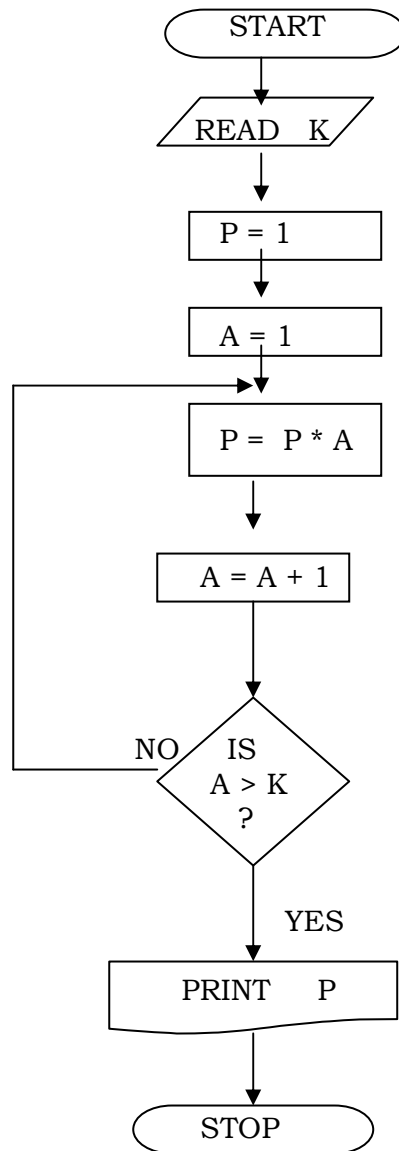


Q3. Draw the program flowchart for finding the sum of squares of first 20 odd numbers

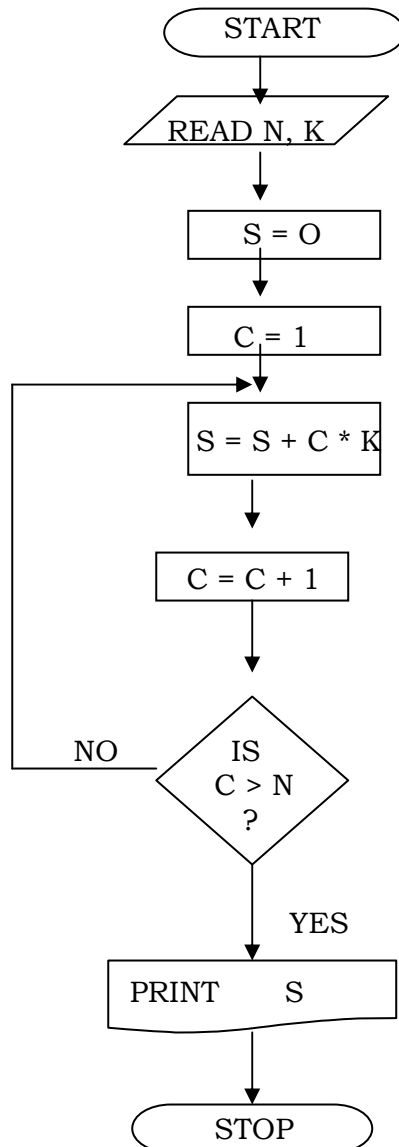


RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

Q4. Draw a flowchart to print the factorial of any given positive non-zero integer K. (The Fctorial of K is defined as: $1 * 2 * 3 * \dots * K$).



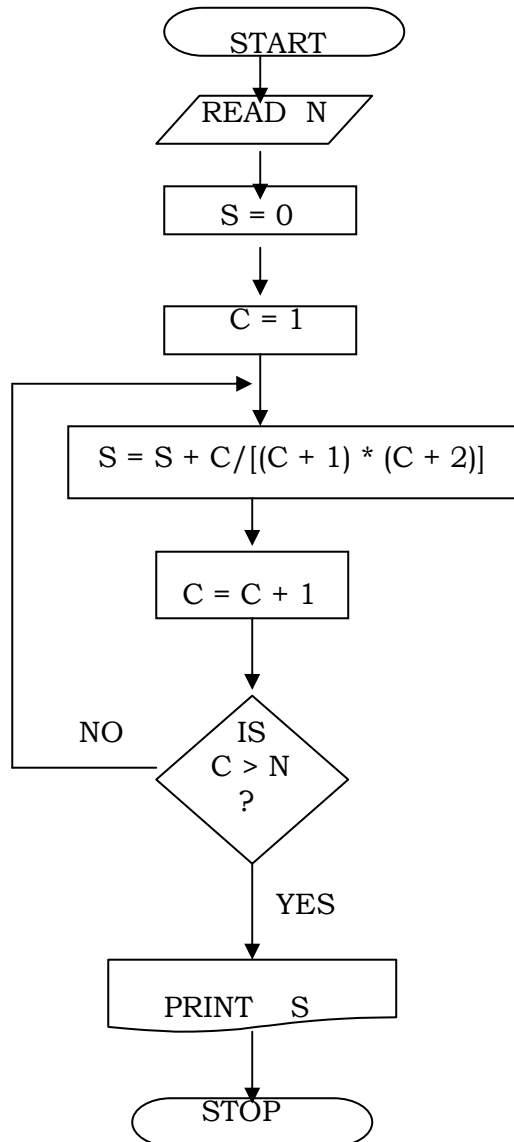
Q5. Draw the flowchart which will calculate the sum of the first N multiples of an integer K, i.e., $1 * K + 2 * K + ... + N * K$



S – Sum of the given series
N – Number of multiples
K – Integer

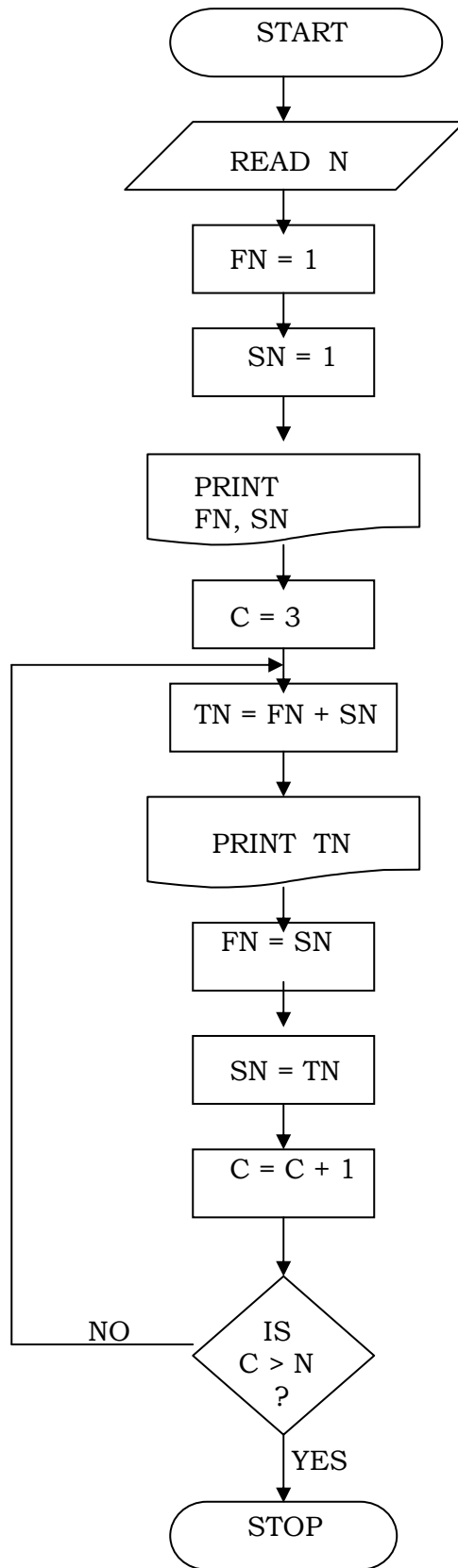
Q6. Draw a flowchart to print the sum of the following series:

$$\frac{1}{2.3} + \frac{2}{3.4} + \frac{3}{4.5} + \dots N \text{ terms.}$$



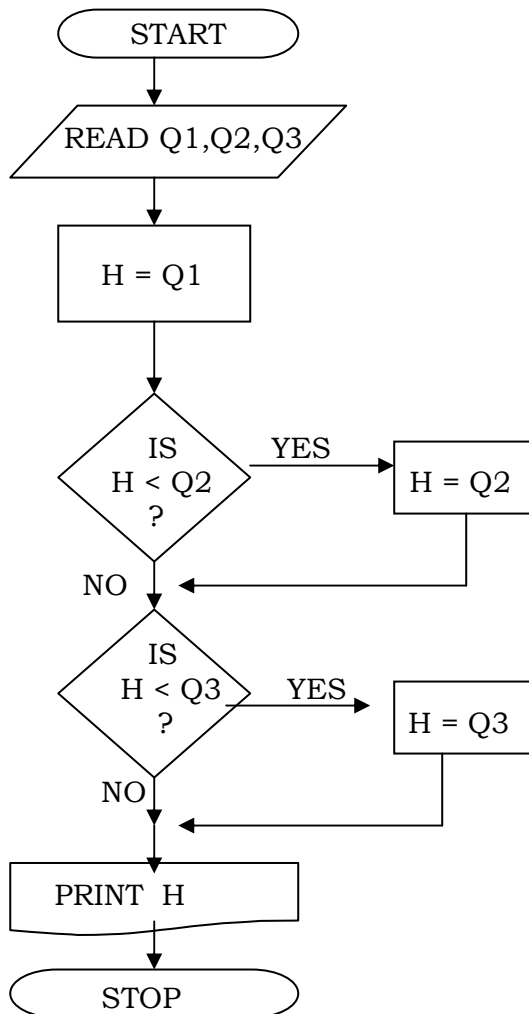
RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

Q7. Draw a flowchart to generate the Fibonacci numbers: 1,1,2,3,5,8,13, ..., N terms (in the Fibonacci series, each number is equal to the sum of the previous two numbers).



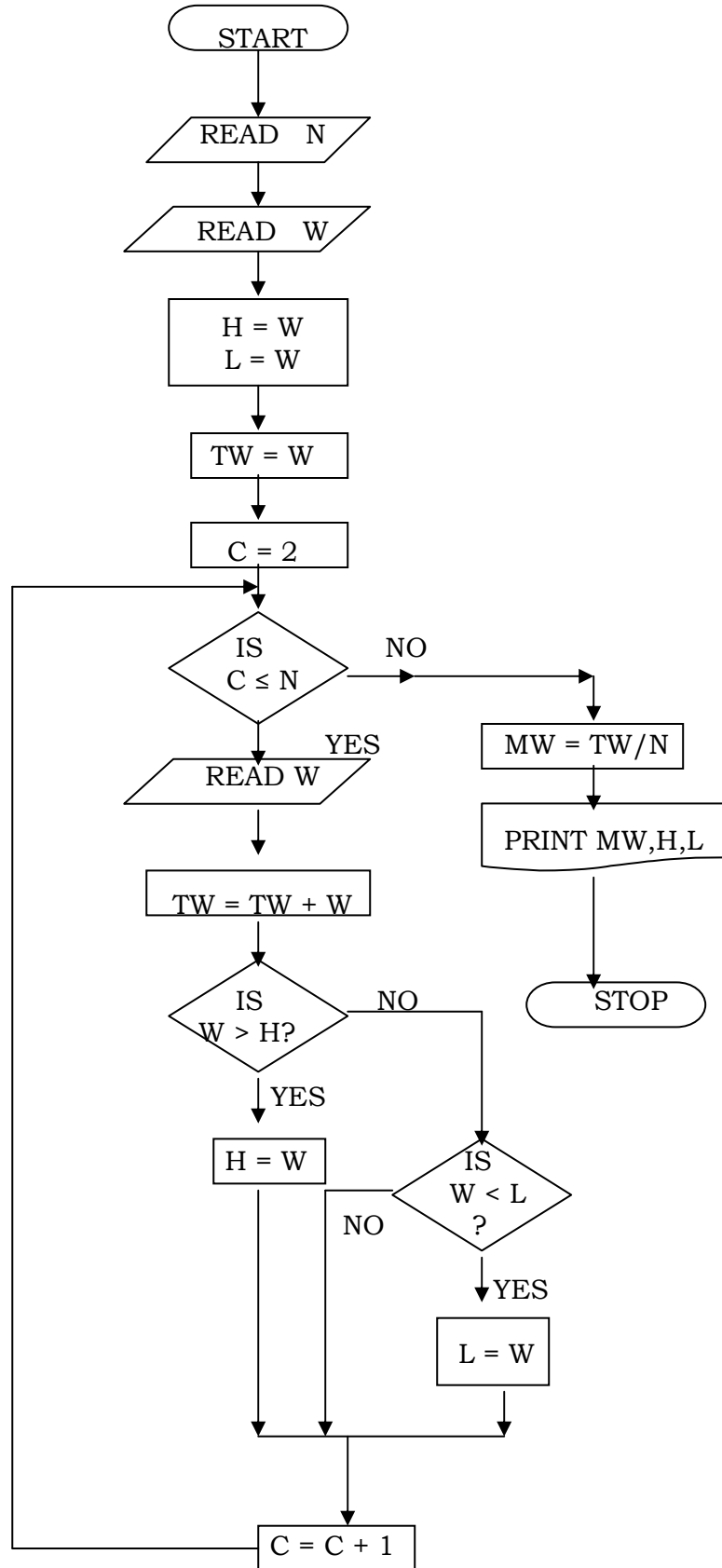
RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

Q8. Draw a program flowchart to find the biggest of three numbers Q1, Q2 and Q3. The biggest is stored in location H.



RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

Q9. The weights of newly born babies in a hospital are input to a computer. The hospital incharge is interested to find the maximum, minimum and mean weights of all the babies. Draw a suitable flowchart

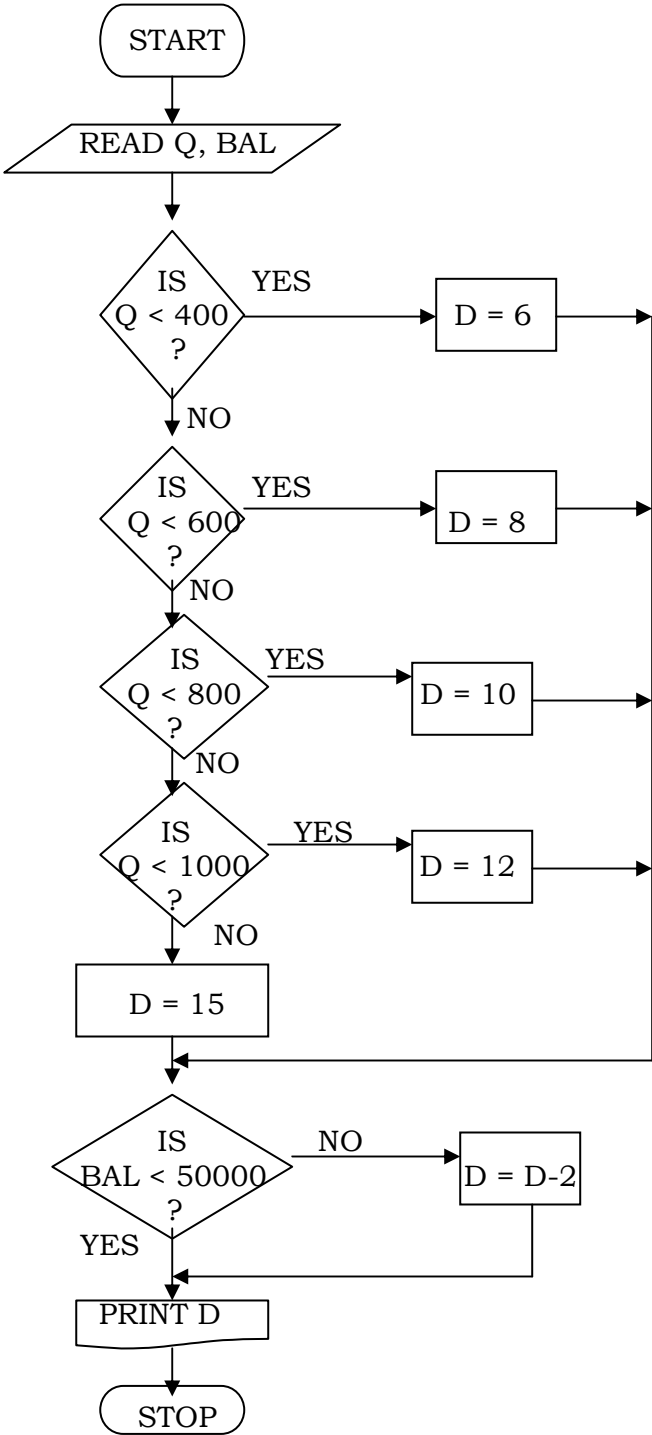


H = HIGHEST WEIGHT
 L = LOWEST WEIGHT
 TW = TOTAL WEIGHT
 N = NUMBER OF BABIES
 W = WEIGHT OF A BOBY

RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

Q10. A company gives discount to its customers on the following basis:
 The adjoining discounts are allowed, if the customer's balance is below Rs.50,000, otherwise the rate of discount is reduced by 2%.

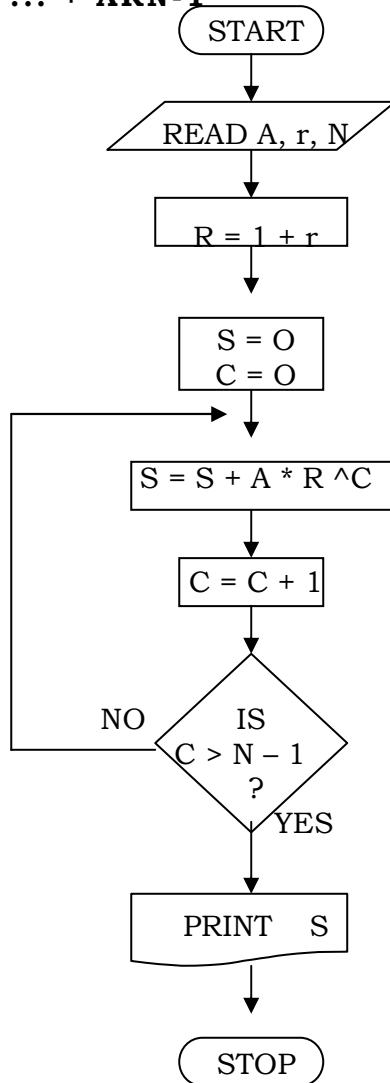
Quantity Ordered	Normal Discount
1 – 399	6%
400 – 599	8%
600 – 799	10%
800 – 999	12%
1,000 and above	15%



RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

Q11. Draw the flowchart for finding the amount of an annuity of Rs. A in N years, rate of interest = r%, $R = 1 + r$ and this amount is given by the following series:

$$S = A + AR + AR^2 + \dots + AR^{N-1}$$

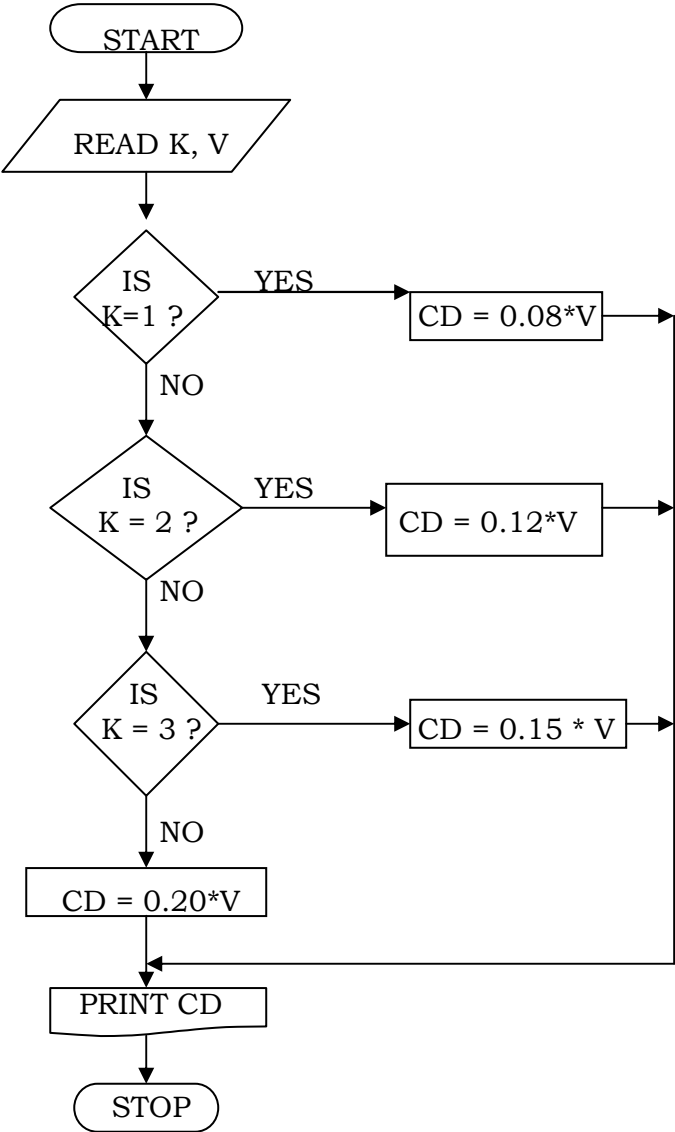


RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

Q12. Assume that imported goods from foreign countries are classified into 4 categories for the purpose of levying customs duty. The rate for each category is as follows:

Class/Categories (K)	Categories of Goods	Custom Duty (%) On Values of Goods (V)
1	Foods, beverages	8
2	Clothing, footwear	12
3	Heavy machinery	15
4	Luxury items	20

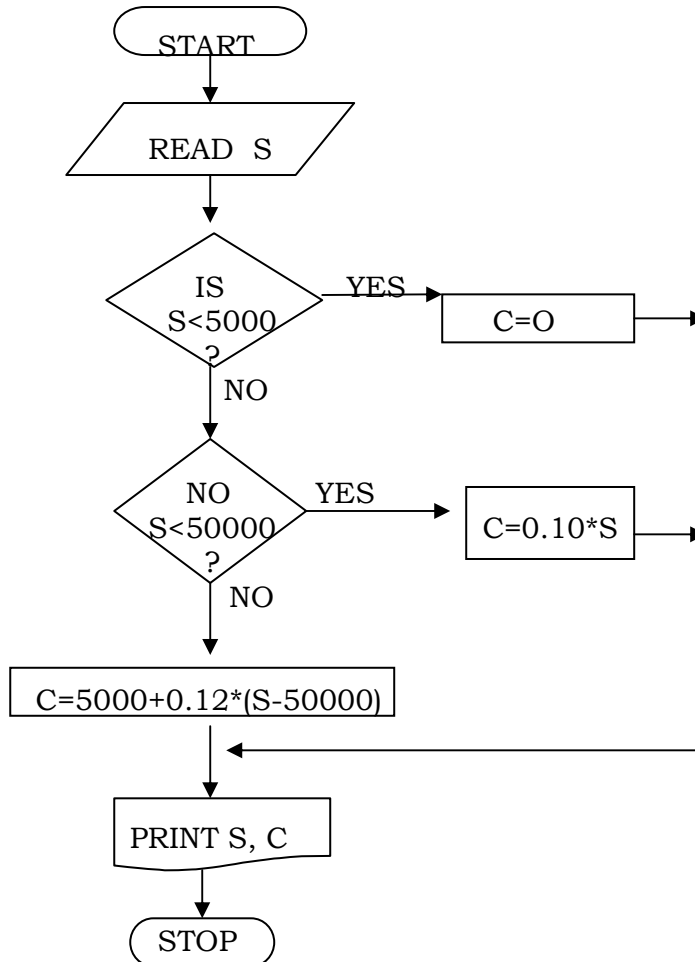
Draw the flowchart for computing the appropriate customs duty.



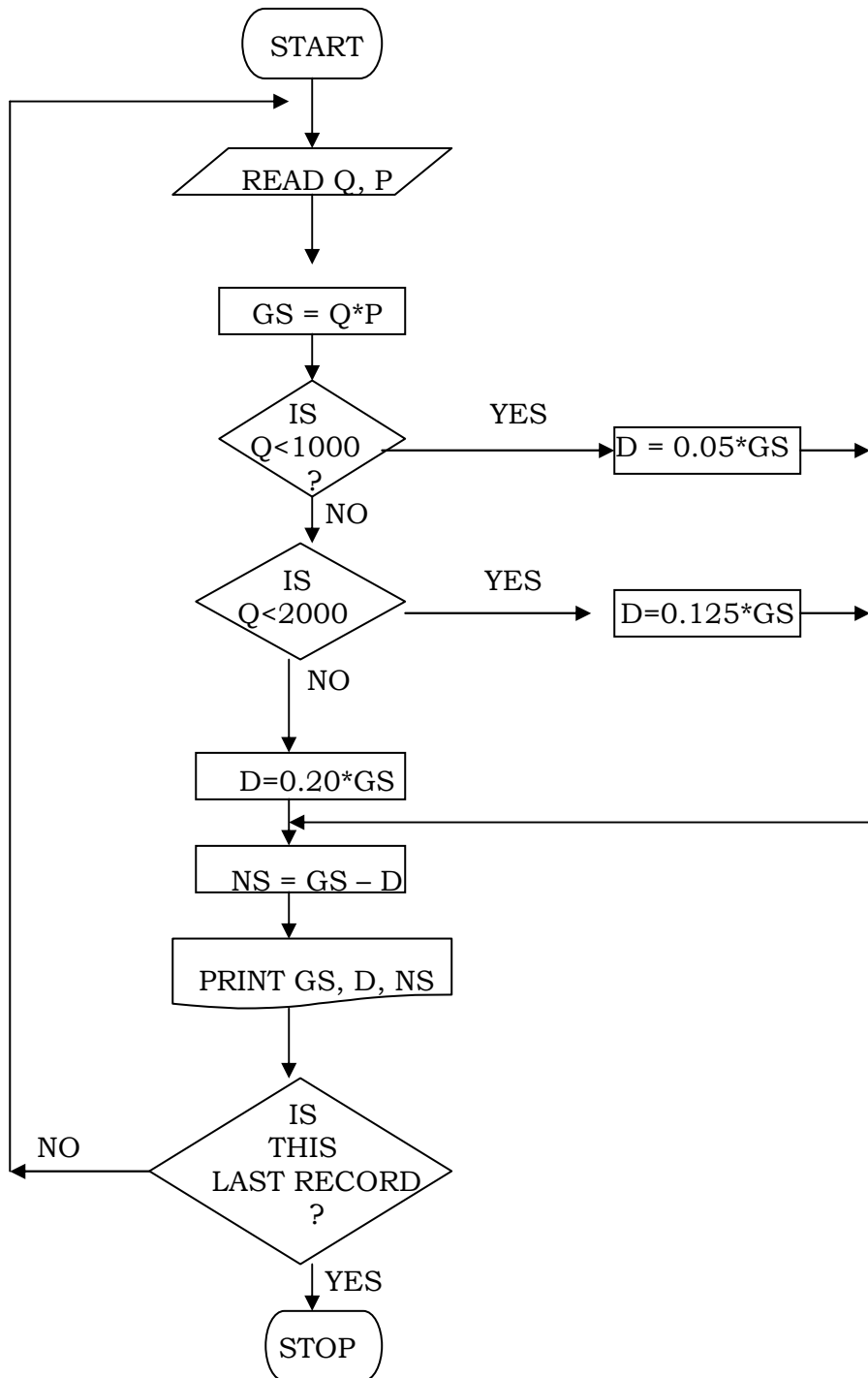
Q13. Draw a flowchart to calculate the commission of a sales representative based on the following set of rules:

- (i) If the sales is less than Rs.5,000, there is no commission.**
- (ii) If the sales is Rs.5,000 or above but less than Rs.50,000, then the commission is computed @10% of the sales.**
- (iii) If the sales is Rs.50,000 or above, the commission is Rs.5,000 and @12% of the sales above Rs.50,000.**

Print the sales and the commission.



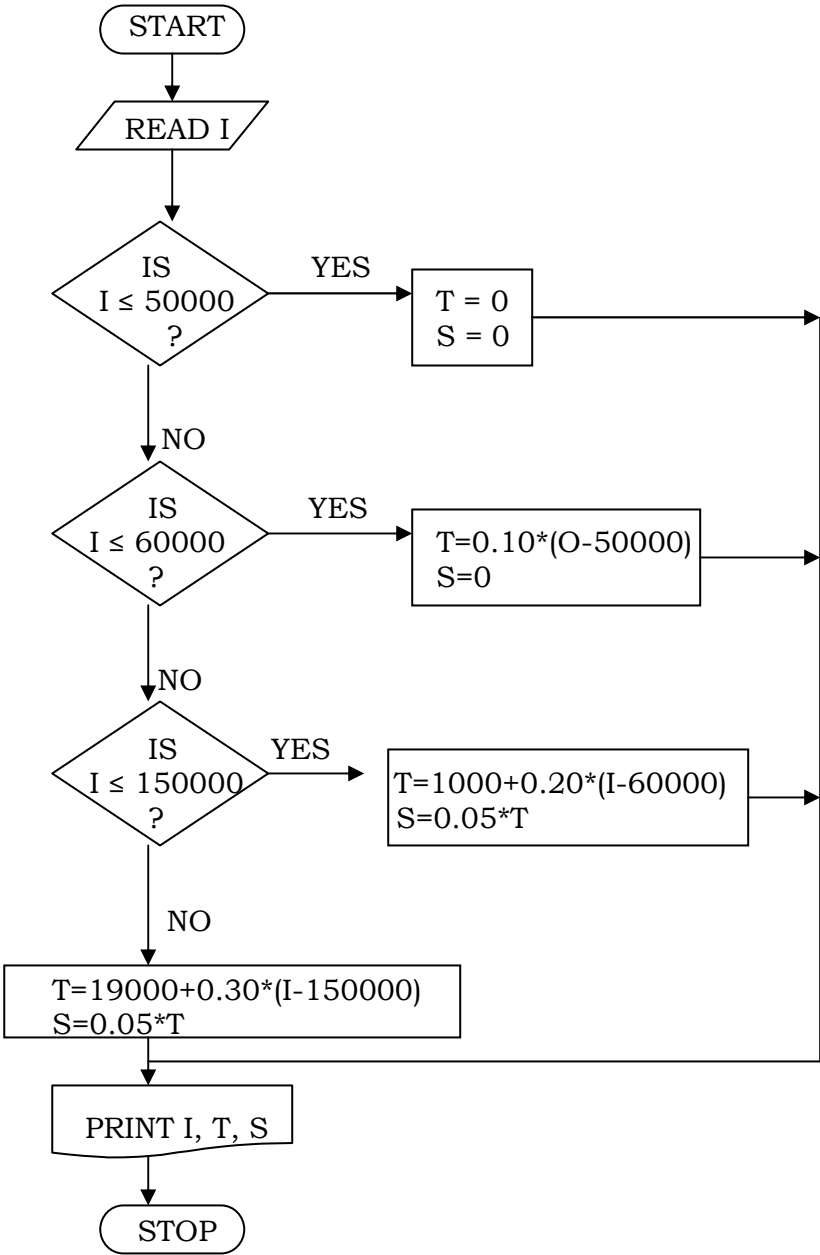
Q14. The problem is to compute, for a series of transactions, the gross sales (G): the quantity discounts, (D), if any; and the net sales (N). The raw data to be supplied in the program includes the quantity sold (Q) and unit price (P). The quantity discount schedule is as follows:
 If quantity sold is: The discount rate would be
 Less than 1,000 units 5%
 1,000 to less than 2,000 12.5%
 2,000 and over 20%



Q15. Draw a flowchart to compute and print the Income-tax and Surcharge on the income of the individual. The income is to be read from the terminal and the tax is to be calculated according to the following rates:

Income (in Rs.)	Rate
Upto 50,000	No tax
From 50,001 to 60,000	10% of the amount exceeds Rs.50,000
From 60,001 to 1,50,000	Rs.1,000 + 20% of the amount above Rs.60,000
Above 1,50,000	Rs.19,000 + 30% of the amount above Rs.1,50,000

The surcharge is levied @5% on the amount of total tax, if the income exceeds Rs.60,000



RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

Q16. Draw a flowchart to illustrate the following situation:

Vishnu Limited calculates discounts allowed to customers on the following basis:

Order Quantity

1 - 99

100 - 199

200 - 499

500 and above

Normal Discount

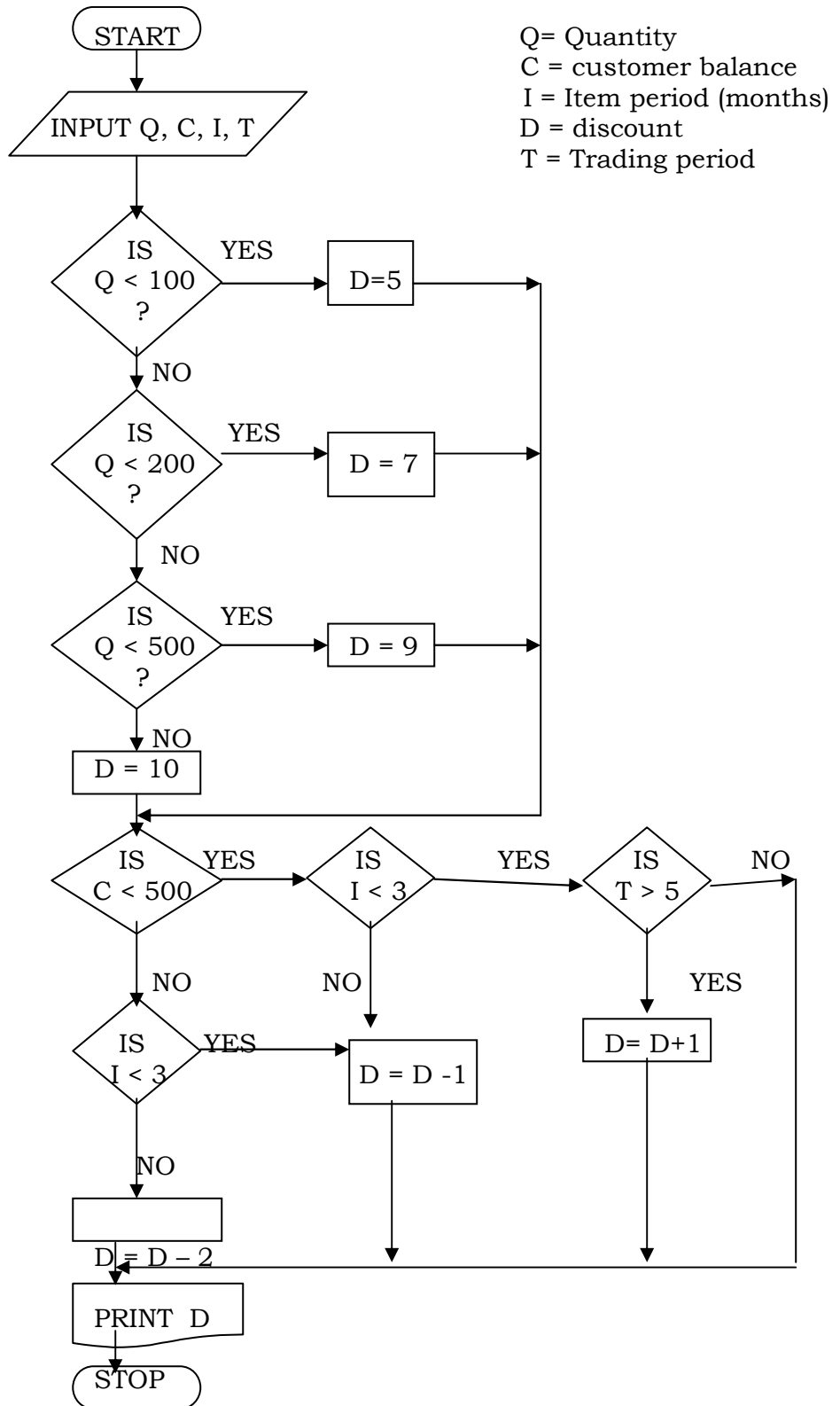
5%

7%

9%

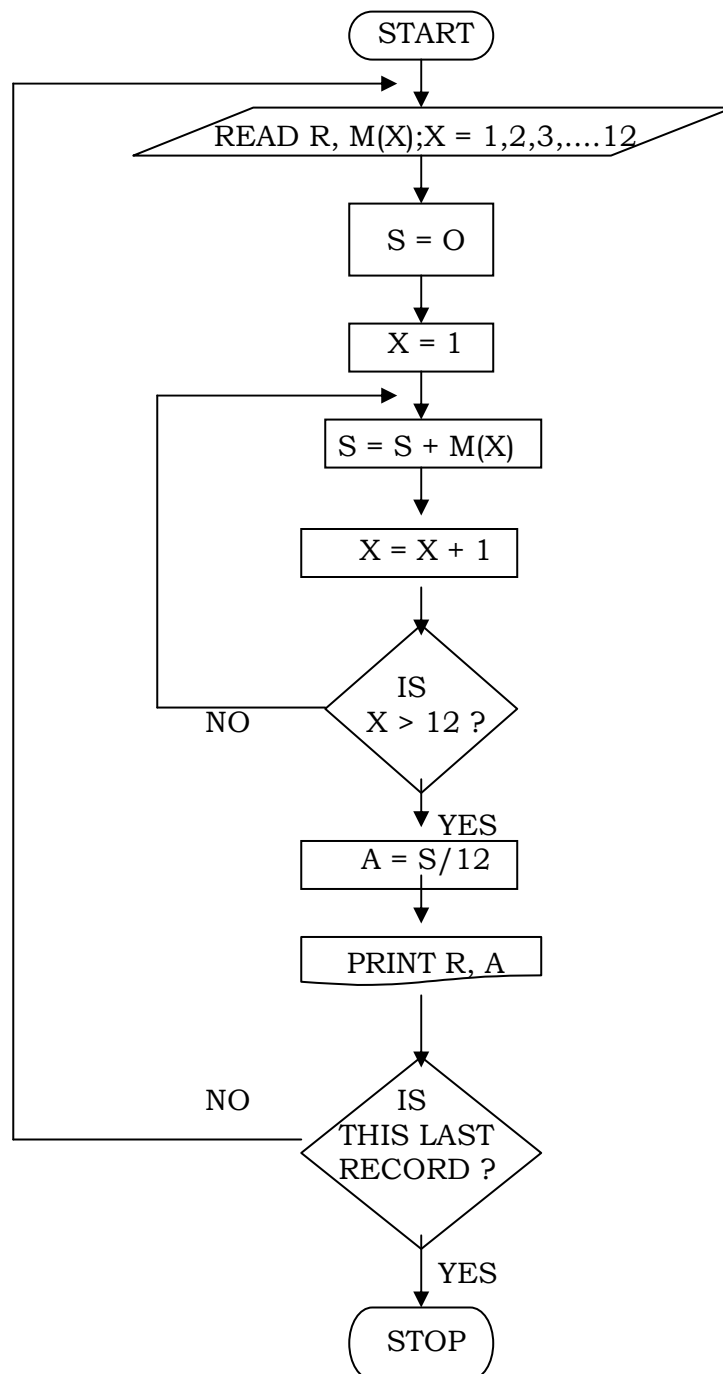
10%

These discounts apply only if the customers account balance is below Rs.500 and does not include any item older than three months. If the account is outside both the limits, the above discounts are reduced by 2%. If only one condition is violated, the discounts are reduced by 1%. If a customer has been trading with Vishnu Limited for over 5 years and conforms to both of the above credit checks, than he is allowed an additional 1%.



RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

Q17. Marks of each student (in a class) in 12 papers are entered through keyboard of a terminal and are read into the CPU locations MARKS 001 to MARKS 012. Draw the flowchart for computing and printing the average marks of each student.



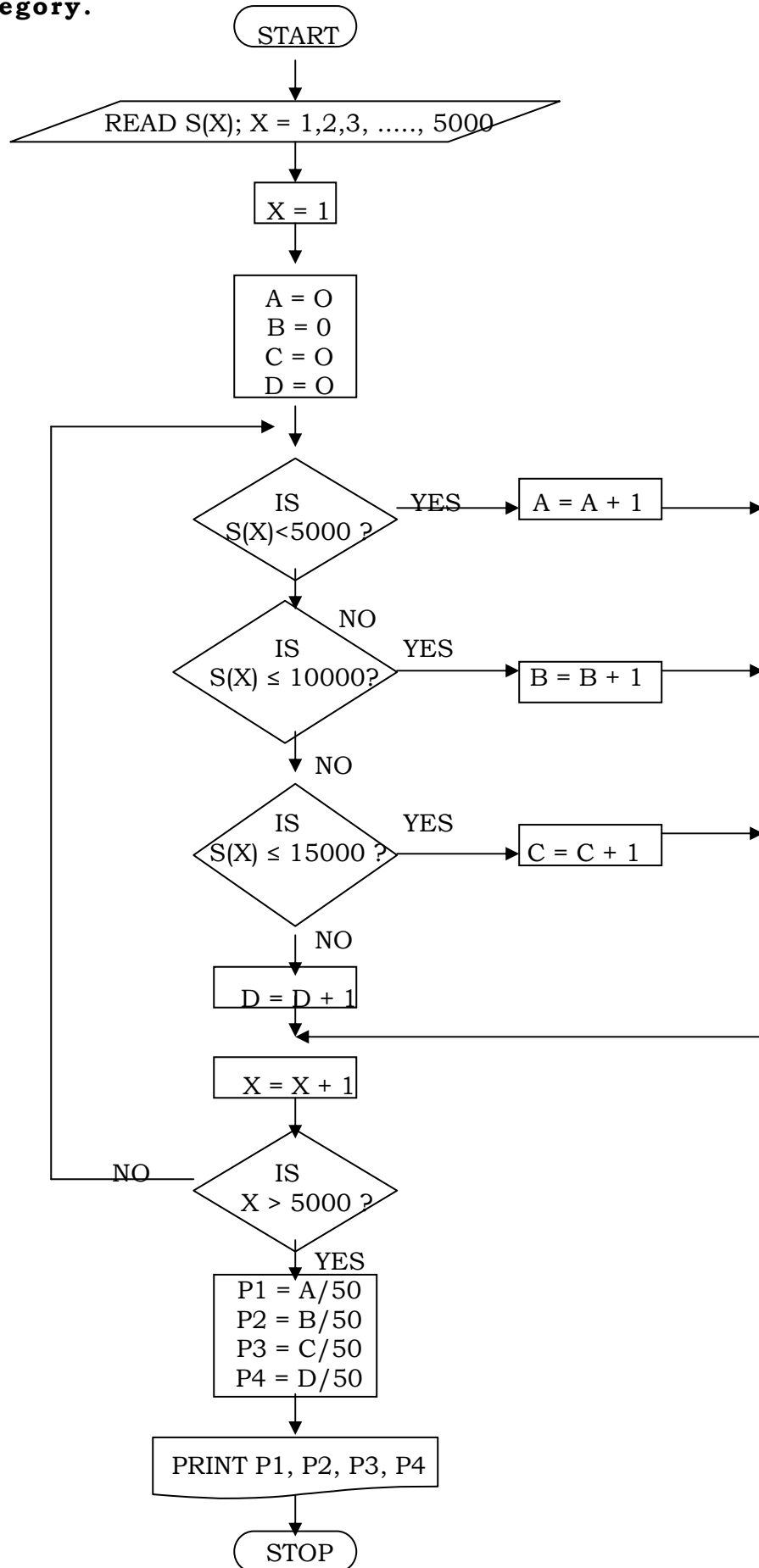
R = ROLL No.
M(X) = Marks in Subject X
S = Total Marks
A = Average Marks

RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

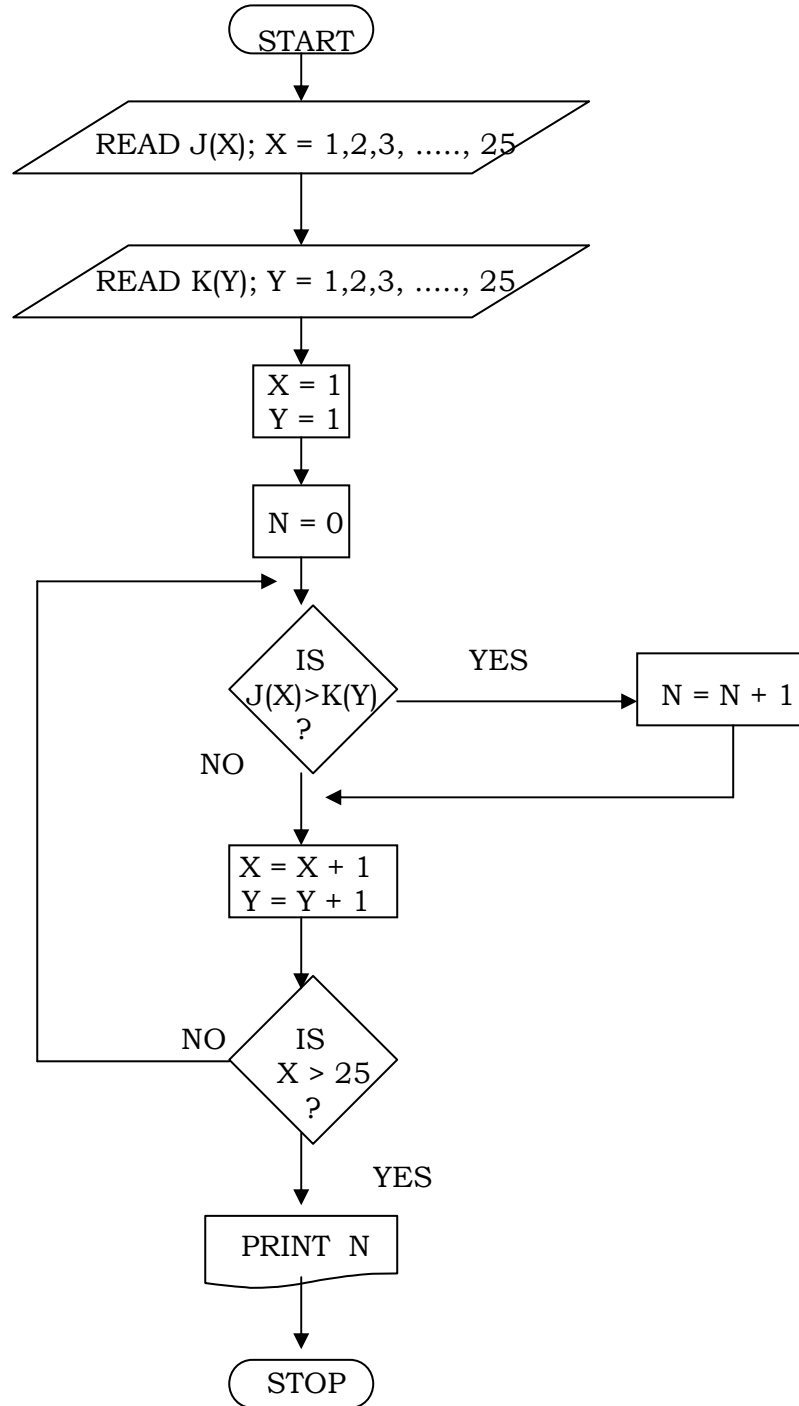
Q18. A company has 5,000 employees. Their salaries are divided in following categories:

- (i) Less than Rs.5,000**
- (ii) Rs.5,000 to Rs.10,000**
- (iii) Rs.10,001 to Rs.15,000**
- (iv) Above Rs.15,000**

Draw the flowchart to find the percentage of the employees in each category.



Q19. Prices for ten commodities in the current year are designated by $J(X)$, X varying from 1 to 25. Likewise, their last year's prices are designated by $K(Y)$, Y varying from 1 to 25. Draw the flowchart for finding the number, N of commodities of which prices have increased.



DECISION TABLES

A decision table is a tabular presentation of program logic. It displays all conditions affecting a particular situation. It gives appropriate action or actions to be taken for each set of conditions.

Decision tables are useful when a computer has to make large number of decisions or if there are a large number of different branches within a program.

Components of a decision table:

Table Heading	Decision Rules Heading
Condition Stub	Condition Entries
Action Stub	Action Entries

Table Heading: The name or number of the Decision Table specifying the problem which is being represented. For complex problems, the decision table is sometimes broken into parts where the Table heading would identify the part of the program being represented.

Condition Stub: The conditions that could exist in the program logic are described.

Action Stub: The action statements or possible outcomes of the system are described.

Decision Rules: These may be one or more decision rules numbered 1,2,3n representing the answers to the Condition Stub and Action Stub described before.

Condition Entry: These represent the possibility of each of the condition stub (answer to each of the condition stub). The condition entry can be 'Y' for Yes or 'N' for No only.

Action Entry: The list relating to the set of action to be performed or taken to a given set of conditions. The action entry can be 'X' for selection action or '-' for unselected action only.

The final decision table would appear as below:

Table Heading	Decision Rule (R1) i.e. a unique combination of 'conditions' and 'actions'	Decision Rule (R2) i.e. a unique combination of 'conditions' and 'actions'	Decision Rule (R.....)
Condition Stub:	Condition Entry i.e. if these conditions exist (Yes)/not exist (No)	Condition Entry i.e. if these conditions exist (Yes)/not exist (No)	Condition Entry i.e. if these conditions exist (Yes)/not exist (No)
C1	Yes/No	Yes/No	Yes/No
C2	Yes/No	Yes/No	Yes/No
C3	Yes/No	Yes/No	Yes/No
.....			
Action Stub:	Action Entry i.e. then the following actions result (X)/not result (-)	Action Entry i.e. then the following actions result (X)/not result (-)	Action Entry i.e. then the following actions result (X)/not result (-)
A1	X / -	X / -	X / -
A2	X / -	X / -	X / -
A3	X / -	X / -	X / -
.....			

RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

Types of Decision Tables:

Limited Entry Decision Table: The Condition Stubs and Action Stubs are exhaustively defined. This means that the condition entry may contain either 'Yes' representing existence of condition or 'No' representing non existence of condition. In the same manner the action entry may contain a 'X' representing execution of a particular action or '-' representing that particular action is not to be executed.

Illustration. A company allows credit to its customers if they are from local city and can either provide surety or have good credit worthiness.

Steps:

- Identify conditions:**
 Local area : Yes/No
 Provide surety : Yes/No
 Good credit worthiness : Yes/No
- Identify Actions:**
 Refuse Credit Grant Credit
- Determine the possible number of rules.** If there are N number of conditions, there are 2^N number of possible rules. In the present case there are 2^3 possibilities. i.e. 8 possibilities
- Draw the table with stubs and required number of entry columns. i.e. column for rules.**
- Fill the condition entries using half rule.**
- For each column enter the corresponding action or actions to be taken.**
- Scrutinise the table to find if any of the rules can be eliminated.**

ALLOWING CREDIT FACILITY	DECISION RULES							
	1	2	3	4	5	6	7	8
CONDITIONS:								
C1: LOCAL CITY	Y	Y	Y	Y	N	N	N	N
C2: PROVIDE SURETY	Y	Y	N	N	Y	Y	N	N
C3: GOOD CREDIT WORTHINESS	Y	N	Y	N	Y	N	Y	N
ACTIONS:								
A1: GRANT CREDIT	X	X	X	-	-	-	-	-
A2: REFUSE CREDIT	-	-	-	X	X	X	X	X

If we observe rules 1 & 2, we find that if C1 & C2 are 'y', it is irrelevant whether C3 is 'y' or 'N' as it does not affect action. So rules 1 & 2 can be combined. Alternatively, rules 1 & 3 can be combined in which case C2 becomes irrelevant.

Again, observe rules 5,6,7 & 8. If C1 is 'N', action is always the same and is not affected by C2 & C3.

Above Table:

R1	R2	R3	R4	R5	R6	R7	R8
New Table		R1	R2	R3	R4		

ALLOW CREDIT FACILITY	DECISION RULES			
	1	2	3	4
CONDITIONS:				
C1: LOCAL CITY	Y	Y	Y	N
C2: PROVIDE SURETY	Y	N	N	-
C3: GOOD CREDIT WORTHINESS	-	Y	N	-
ACTIONS:				
A1: GRANT CREDIT	X	X	-	-
A2: REFUSE CREDIT	-	-	X	X

RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

Extended Entry Decision Table:

This is second type of decision table. Here, the condition and action stub are generalised. The statements made in the stub portion are incomplete. The condition and action entry are of descriptive type.

Illustration: A life insurance company has the following criterion for deciding whether or not the insurance for a person is to be accepted.

If a person's health is good and the person is aged between 20 and 35 years, lives in a town or city, and is a male, then the premium charged is Rs.20 per thousand and the policy is written for an amount not exceeding Rs.10,00,000. If a person satisfies all the above conditions except that the person is a female, then the premium charged is Rs.25 per thousand and the policy is written for not more than Rs.6,00,000. If the person's health is poor and the person is aged between 20 and 35 years, lives in a village and is a male, then the premium charged is Rs.40 per thousand and the policy is written for not more than Rs.2,00,000. If the person is a female, the premium rate charged and the maximum limit for writing a policy are the same. In all other cases, the person is refused insurance.

Prepare a decision table for the above problem.

LIMITED ENTRY DECISION TABLE

POLICY OF INSURANCE COMPANY	DECISION RULES			
	1	2	3	4
CONDITIONS :				
C1: Is health good?	Y	Y	N	E
C2: Is age between 20 and 35 years?	Y	Y	Y	L
C3: Is the person male?	Y	N	--	S
C4: Does the person live in town or city?	Y	Y	N	E
ACTIONS				
A1: Premium Rs.20 per thousand	X	-	-	-
A2: Maximum policy amount Rs.10,00,000	X	-	-	-
A3: Premium Rs.25 per thousand	-	X	-	-
A4: Maximum policy amount Rs.6,00,000	-	X	-	-
A5: Premium Rs.40 per thousand	-	-	X	-
A6: Maximum policy amount Rs.2,00,000	-	-	X	-
A7: Refuse insurance	-	-	-	X

EXTENDED ENTRY DECISION TABLE

POLICY OF INSURANCE COMPANY	DECISION RULES			
	1	2	3	4
CONDITIONS :				
C1: Health	Good	Good	Poor	E
C2: Age	20 to 35	20 to 35	20 to 35	L
C3: Sex	Male	Female	--	S
C4: Place of Residence	Town/city	Town/city	Village	E
ACTIONS				
A1: Premium per thousand	Rs.20	Rs.25	Rs.40	-
A2: Maximum policy amount	Rs.10,00,000	Rs.6,00,000	Rs.2,00,000	-
A3: Insurance Policy	ALLOW	ALLOW	ALLOW	REFUSE

RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

Mixed entry decision table

This is third type of decision table. It uses the techniques of both a limited entry decision table and extended entry decision table. While the limited and extended entry forms can be mixed within the table, only one form may be used within a condition statement/entry or an action statement/entry.

Illustration: A wholeseller has three commodities to sell and has three types of customers. The discount is given as per following rules:

- (i) For Govt. orders, 15% discount is given irrespective of the value of the order.
- (ii) For orders of more than Rs.20,000, an agent gets a discount of 20% and the retailer 15% respectively.
- (iii) For orders of value between Rs.10,000 and Rs.20,000, agent gets discount of 15% and the retailer gets 10%.
- (iv) For order of value less than Rs.10,000, the agent and retailer get discount of 10% and 5% respectively.

The above rules do not apply to furniture items. However, in case of furniture items, a flat rate of 10% discount is admissible to all type of customers.

Prepare a mixed entry decision table.

MIXED ENTRY DECISION TABLE

Discount policy	Decision Rules							
	1	2	3	4	5	6	7	8
Conditions								
C1: Order value	Any amount	Any amount	More than Rs.20,000		Between Rs.10,000 & Rs.20,000		Below Rs.10,000	
C2: customer type	Any	Govt.	Agent	Retailer	Agent	Retailer	Agent	Retailer
C3: Product	Furniture	Others	Others	Others	Others	Others	Others	Others
Actions								
Discount								
A1: 5%								X
A2: 10%	X					X	X	
A3: 15%		X		X	X			
A4: 20%			X					

Benefits of Decision Tables:

Following are given some of the Benefits of Decision Tables:

1. **Easy to Draw** - Decision Tables are easy to draw and modify as compared to flowcharts.
2. **Compact Documentation** - The documentation in the form of decision tables is compact since one decision table may replace few pages of a flowchart
3. **Simplicity** - It is easier to follow a particular path in one column of a decision table than it is to go through several pages of the flowcharts.
4. **Direct Codification** - The decision tables can be directly coded into a program.
5. **Better Analysis** - A decision table shows various alternatives and their respective outcomes side by side for better analysis of the problem.
6. **Modularity** - The complex problems would require complex decision tables which can be easily broken down to micro-decision tables.
7. **Non-technical** - No knowledge of computer language or CPU working is necessary for drawing decision tables.

Limitations of Decision Tables

Following are given some of the Limitations of Decision Tables:

1. All programmers may not be familiar with Decision Tables and therefore flow charts are more common

RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

2. Flowcharts can better represent a simple logic of the system rather than a decision table.
3. The decision tables do not express the total sequence of the events needed to solve the problem.

SOLVED ILLUSTRATIONS:

1. If a customer has good credit worthiness and in addition either he has no outstanding loan or he can provide a surety, he can be granted a loan by the Bank.

Scrutiny for Loan	1	2	3	4
Conditions				
Good Credit Worthiness	Y	Y	Y	N
Outstanding Loan	Y	Y	N	-
Provide Surety	Y	N	-	-
Actions:				
Give Loan	X	-	X	-
Refuse Loan	-	X	-	X

2. While invoicing each customer, the clerk has to work out the discount allowable on each order. Any order above Rs.20,000 attracts a bulk discount of 8%. A customer within the trade is allowed 10%. There is also a special discount of 5% allowed for any customer who has been ordering regularly for over 5 years.

Discount Processing	1	2	3	4	5	6	7	8
Conditions								
Order value-more than Rs.20,000	Y	Y	Y	Y	N	N	N	N
Customer within the trade	Y	Y	N	N	Y	Y	N	N
Customer ordering regularly for over 5 year	Y	N	Y	N	Y	N	Y	N
Actions:								
Discount - NIL								X
Discount - 5%							X	
Discount 8%				X				
Discount - 10%						X		
Discount -13%			X					
Discount - 15%					X			
Discount-18%		X						
Discount-23%	X							

3. A company offers hire-purchase finance where the customers get the credit facility if they satisfy any of the following conditions:-
 - a. The customer must hold the present job for more than 5 years and reside in the same place for atleast 3 years. In such case he would get credit upto Rs.10000/-
 - b. The monthly salary of customer must exceed Rs.5000 and must hold the present job for more than 5 years. In this case the credit would be given upto Rs.15000.
 - c. The monthly salary exceeds Rs.5000 and reside at the same place atleast for 3 years, he is allowed a credit of Rs.20,000.
 - d. In case the customer's monthly salary exceeds Rs.5000 and he holds the present job for more than 5 years and also he resides at the same place atleast for 3 years, the credit facility is allowed upto Rs.25,000.
 - e. For all other customers, request for credit facility is rejected.

RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

Hire-Purchase Facility	1	2	3	4	5	6	7
Conditions:							
Customer holds job for more than 5 years	Y	Y	Y	Y	N	N	N
Reside at same place for atleast 3 years	Y	Y	N	N	Y	Y	N
Monthly salary exceed Rs.5000	Y	N	Y	N	Y	N	-
Actions:							
Give Credit-Rs.10,000		X					
Give Credit-Rs.15000			X				
Give Credit-Rs.20,000					X		
Give Credit-Rs.25000	X						
Refuse Credit				X		X	X

4. The numerical scores in an examination (0 to 100) are to be converted into letter grades (A to D) using the following procedure:-

Numerical Score	Letter Grade
Less than 40	D
40 to 55	C
55 to 85	B
More than 85	A

Students Grading	1	2	3	4
Conditions:				
Less than 40	Y	N	N	N
40 to 55	-	Y	N	N
55 to 85	-	-	Y	N
Actions:				
Letter grade-A				X
Letter grade-B			X	
Letter grade-C		X		
Letter grade-D	X			

5. A firm keeps details of the bills raised on customers in a computer file. The record has following data fields:

Customer name, Type (Dealer/Other), Bill number, Bill data, Amount, Date of Payment.

If a Customer pays the bill within one month, he is allowed 10% discount. If he pays within 1 to 2 months, no discount is allowed. If he pays after 2 months, penal interest 10% as charged from him. If the customer is a dealer, then the corresponding discount & interest rates are 15%, 0, 10% respectively.

Discount/interest processing	1	2	3	4
Conditions				
Customer is a dealer	Y	-	N	-
Pays the bill within one month	Y	N	Y	N
Pays bill in 1-2 months	-	N	-	Y
Actions:				
Discount- Nil				X
Discount-10%			X	
Discount-15%	X			
Penal Interest-10%		X		

6. A Company provides home delivery for its products and bills delivery charges according to the distance and the amount of the bill. If the distance of customer and company's office is less than 5 KM, no delivery charges are levied. If the distance is more than 5KM but less than 15KM, the delivery charges are 2% of the bill amount. However, if the bill amount is more than Rs.1000 then a concessional rate of 1% is levied. If

RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

the distance is more than 15KM, the delivery charges are 5% but if the bill amount is more than Rs.1000/- then 2% delivery charges are levied.

Delivery Charges processing	1	2	3	4	5
Conditions:					
Distance-Less than 5KM	Y	N	N	N	N
Distance-More than 5KM but less than 15KM	-	Y	Y	N	N
Bill Amount-Upto Rs.1000	-	Y	N	Y	N
Actions:					
Delivery Charges-NIL	X	-	-	-	-
Delivery Charges-1%	-	-	X	-	-
Delivery Charges-2%	-	X	-	-	X
Delivery Charges-5%	-	-	-	X	-

7. A Finance Company provides loans for building residential houses or for purchasing vehicles and other kind of loan applications are rejected straightaway. The home loan or vehicle finance applicants must be serving in Govt. or Public Limited Company and have good credit rating. Father, all home loan applicants must have completed atleast 5 years of regular service, have a balance service of atleast 15 years and have a life insurance policy backup. The vehicle loan applicants must have completed a service of atleast 3 years, have a balance service of atleast. 7 years and must have a valid driving licence.

Preliminary Scrutiny of Loan applications	1	2	3	4
Conditions:				
Service in Govt. or Public Limited Company	Y	Y	Y	N
Good Credit Rating	Y	Y	N	-
Application for Home Loan	Y	N	-	-
Actions:				
Go to Home Loan Table	X	-	-	-
Go to Vehicle Loan Table	-	X	-	-
Reject application	-	-	X	X

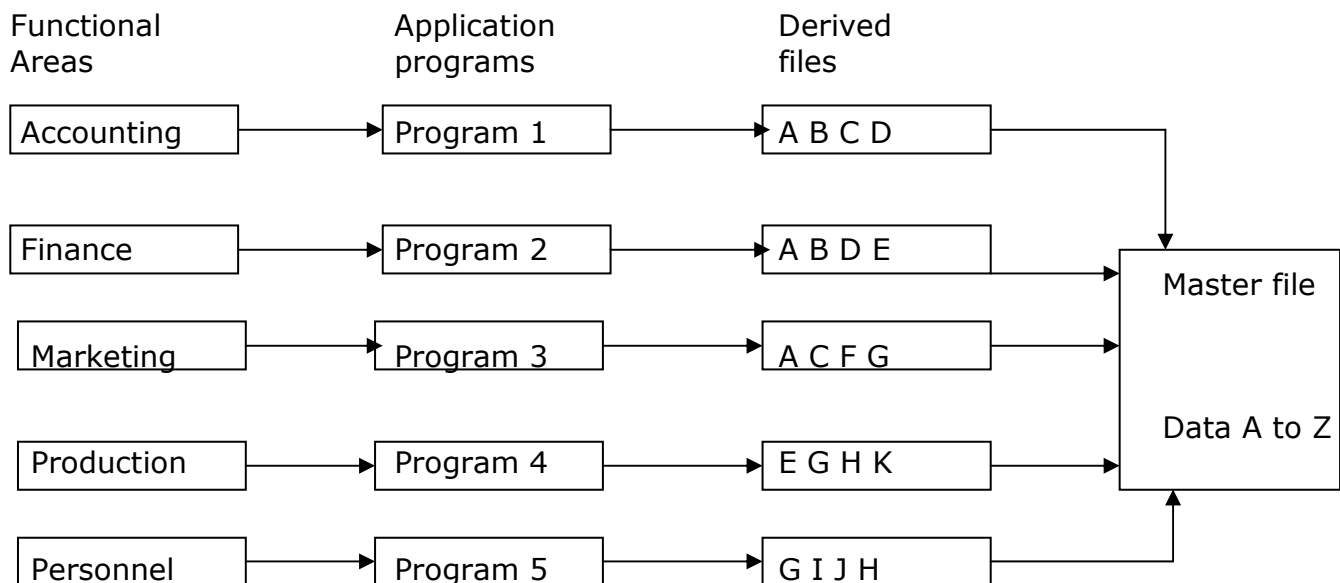
Home Loan Scrutiny	1	2	3	4
Conditions:				
Completed minimum 5 years of service	Y	Y	Y	N
Balance service of minimum 15 years	Y	Y	N	-
Have a Life Insurance Policy backup	Y	N	-	-
Actions:				
Accept application	X	-	-	-
Reject application	-	X	X	X

Vehicle Loan Scrutiny	1	2	3	4
Conditions:				
Completed minimum 3 years of service	Y	Y	Y	N
Balance service of minimum 7 years	Y	Y	N	-
Have a valid driving licence	Y	N	-	-
Actions:				
Accept application	X	-	-	-
Reject application	-	X	X	X

DATA BASE MANAGEMENT SYSTEM

PROBLEMS IN TRADITIONAL FILE MANAGEMENT

In the past, traditional file management had been the outcome of information processing on a small scale, automating one application at a time. Information systems tended to grow independently in individual organisation. Thus, each division of an organisation having multiple functions developed its own applications. Within each division, each functional area tended to develop systems in isolation from other functional areas. This practice led to the emergence of information processing as shown in fig. below:



Thus, in traditional file management, each functional area develops its own specialised applications. Data needed for each user application was stored in independent data files. Processing consisted of using separate computer programs that updated these independent data files and used them to produce the documents and reports required by each separate user application. This file processing approach is still being used, but it has several problems which are discussed below that limit its efficiency and effectiveness for end user applications.

- 1. Data Duplication i.e., data redundancy:** Data redundancy occurs because different divisions, functional areas, and other groups of the organisation collect the same piece of information independently. This data redundancy causes problems when data has to be updated, since separate file maintenance programs have to be developed and coordinated to ensure that each file is properly updated.
- 2. Data Dependence:** In file processing systems, major components of a system i.e., the organization of files, their physical locations on storage, hardware and the application software used to access those files depend on one another in significant ways. If changes are made in the format and structure of data and records in a file, changes have to be made in all the programs that use this file. This program maintenance effort is a major burden of file processing systems. It is difficult to do it properly, and it results in a lot of inconsistency in the data files.
- 3. Lack of Data Integration:** Having data in independent files makes it difficult to provide end users with information for ad hoc requests that require accessing data stored in several different files. Special computer programs have to be written to retrieve data from each independent file. This is difficult, time-consuming, and expensive for the organizations.
- 4. Lack of Security:** Traditional file management lacks security as there is little control over the access to and dissemination of information. As the access is to the whole file and not only a part of it, whole data contained in a file can be accessed. The restriction to data access is at the most in the form of difficulty in locating the information.
- 5. Data Isolation:** The data is scattered in various files at different locations. The files also may be stored in the different formats. It is then difficult to write new application program that will retrieve the appropriate data.

6. **Lack of Data Sharing:** Data cannot be used by multi users. This happens because pieces of information in different files and in different parts of the organisation cannot be related to one another.
7. **Lack of Flexibility:** Traditional file management lacks flexibility as it can deliver only routine scheduled reports but it cannot deliver ad hoc reports or respond to unanticipated information requirements in timely fashion.
8. **Integrity Problem:** The data stored should satisfy some conditions before stored in the files. These are known as 'Consistency constraints'. Developers are responsible for enforcement by applying these rules in their programs. Whenever the new constraints are added, it is difficult to change the programs to enforce them. The problem gets intensified when constraints involve several data items from different files.
9. **Concurrency Problems:** In many applications, it is very essential that multiple users access and update the same data simultaneously. The concurrent updates may result in inconsistent data. System should have some supervision mechanism in the concurrency control. This is also very difficult in the file processing system.

DATABASE

"A 'Database' is a computer file system that uses a particular file organization to facilitate rapid updating of individual records, simultaneous updating of related records, easy access to all records, by all applications programs, and rapid access to all stored data which much be brought together for a particular routine report or inquiry or a special-purpose report or inquiry." - G.M. Scott

In order to overcome the problems of a traditional file-oriented database processing system, the database concept was introduced. It is defined as an integrated collection of data and information which is utilized by several information sub-systems of an organisation. Such a database can be organized as an integrated collection of data records into a single 'superfine' or it can be organized as an integrated collection of several data files. The organization of a common database allows it to be accessed by several information sub-systems and thus eliminates the necessity of duplication in data storage, updating deletion and protection.

Regardless of its file organisation, a database system includes several components that collectively give it following distinct characteristics:

1. Centralised and integrated shared data file which consists of all data used by a company.
2. Organized and structured in a different manner than the conventional sequential file organizations.
3. Organization is such that duplication of data is minimized if not eliminated entirely.
4. Permits access to any or all data quantities by all applications with equal ease.
5. Emphasises the independence of programs as well as data and involves the concept of separating data definition from the applications programs and including it as part of the data base.
6. Stored on a direct access storage device.
7. Provides for the definition of logical relationships which exist between various records in the database.
8. Acts as common data source to various users and thus avoids duplication in storage and retrieval of data and information.
9. Is controlled by a separate authority established for the purpose.
10. A database is a collection of related files, and offers users more flexibility than files. With a database data are independent of the application programs that use them; data are accessible to any program regardless of the location of the data or the language in which the program is written; data are not duplicated in different locations, and programmers need not write and debug extensive file descriptions in order to work with data.

THE MAIN FUNCTIONS OF DATABASE ARE AS FOLLOWS:

- (i) Databases reduce the data redundancy to a larger extent. Data redundancy means duplication of data. Non-database system maintains a separate copy of data for each application. For example, in a school environment, a separate file is maintained for fees and bus information. Suppose the permanent address of a student gets changed. The student informs about it to the bus authorities and the changed address is reflected in

the students' record. But this change is not reflected in the fees database. Now we have two different records for same student in two different databases. This may lead to inconsistent data.

- (ii) Databases can control data inconsistency to a larger extent. When the redundancy is not controlled, there may be occasion that two entries about the same data do not agree. By controlling the redundancy, inconsistency is also controlled. The problem of inconsistency is removed if the database is centrally administered.
- (iii) Databases facilitate sharing of data. In a centralized database, the basic data is stored in one format but the way of interpretation of data for different users is different. We can say that different users can share the same data visualizing it in their own way.
- (iv) Databases enforce standards. The centralized control of the database can apply certain standards in data representation. These standards may be company's own standards or they may be national or international standards.
- (v) Databases can ensure data security. The database administrator has the complete control over the database. He/She can ensure that access to database is only through permission of the authority. The users of the database can be divided into different categories. Some of the users may be given authority to read and use the data. Only selected users may be given authority to update data or delete records.
- (vi) Databases can maintain integrity. Suppose a database contains information about the marks scored by a set of students in an examination. The maximum marks are 100. The marks scored by a student cannot exceed 100. Integrity checks can be inbuilt into the system so that database will accept numbers only in the range of 0 to 100 as the marks scored by the student.

ARCHITECTURE OF DATA BASE:

Data base follows a three level architecture i.e., it may be viewed from three perspectives:

a. The external view or user view:

It consists identification and description of each data element (field) needed for a given application. It is user view. It includes only those portions of data base or application programs which are of concern to the users. Example one user may want to view record indicating student roll no., name and address while other user may want to view record indicating student roll no. and marks in 6 subjects.

b. Conceptual or Global view:

It is a logical view of all data base entities and their relationship. It is a single view of the entire data base. It describes all records, relationship and constraints. Example conceptual view may define student roll no., name and address as string and marks as integer.

c. Internal view or physical view:

This view indicates the way data will be stored. It indicates data structure and access methods. Example, internal view may define name as comprising 20 characters, marks in each subject as comprising 3 numbers.

ADVANTAGES OF A DATA BASE INCLUDE:

1. Collection of files grouped together as one entity. This collection is tied together by a series of tables and arrays to form one giant integrated file. The tables and arrays serve as an index to define relationships between records and files maintained in the database. In this manner, a transaction in one file of the database can also trigger a series of updates in parts of other files. Thus, data is input only once to the database and is made available to the many files composing it.
2. Use of a database avoids the duplication of files, the reprocessing of the same data items, and the repeated handling of data.
3. The savings attributed to the use of a database result from the computer sharing of records, reduced processing times, reduction in the use of software and hardware, more efficient use of data processing personnel, and an overall improvement in the flow of data.
4. Information supplied to managers is more valuable because it is based on a comprehensive collective of data instead of files which contain only the data needed for one application.
5. Use of integrated systems is greatly facilitated.

6. Great deal of programming time is saved because the DBMS handles the construction and processing of files as well as retrieval of data.
7. The database also maintains an extensive inventory control file. This file accounts for all parts and equipment throughout the maintenance system. It defines the status of each part and its location.
8. One of the benefits of the database is the timely and accurate reporting of data to all maintenance centres.

DISADVANTAGES OF A DATABASE ARE:

1. Its size, its cost, and the time required to implement such a system. A database requires the use of a large-scale computer system.
2. A large full-time staff is also required to design, program, and support the implementation of a database.
3. The cost of a data base project is a prohibitive factor for many organizations. Database-oriented computer systems are not luxuries and are undertaken when proven economically justifiable.

DATA DICTIONARY:

Data dictionary, sometimes also called as 'data directory', is a tool for organizing and storing information about the data maintained in the database. It is a file that contains meta-data i.e., data about data. This file is consulted before actual data is read or modified in the database system. A data dictionary contains the following information about data:

1. Name of the data item.
 2. A description of the data item- language description, processing description, allowable values, edit and validation criteria, calculations to determine value, classification, security, etc.
 3. Source of data – Various sources of data input.
 4. Impact analysis – users of the data including screens, reports, programs, and organizational positions that use the data item.
 5. Key words used for categorizing and searching for data item descriptions.
- Usually, there are three types of facilities in a data dictionary:
- (i) Report facility: detailed reporting on data items and their uses, cross-reference reports showing relationships, and summary reporting providing various summary data.
 - (ii) Control facility: detects violations of standards for user authority, documentation, etc.
 - (iii) Excerpt facility: specific task performance, such as preparation of test data, copying of code from existing programs and inserting in programs under development, and copying documentation into source code.

The data dictionary provides lists of data items sequenced alphabetically by classification, keyword, etc. The main advantage of a data dictionary is to provide consistency of data descriptions and naming. It also provides ease of updating where one description serves many purposes. For example, if a change is to be made in product code, the data dictionary can identify every program, input screen, and report in which the product code appears.

PROBLEMS OF CREATING A DATABASE:

The problems which are usually observed during the stage of creating a common database in an existing organization are as follows:

- (i) **Maintenance of Database security:** 'Privacy' and 'Integrity' are two aspects which are closely related with the security of common database. 'Privacy' means that data can be accessed by authorized persons only. Integrity means that the changes in the programs and data can be affected only by authorized persons.
- (ii) **Maintenance of Data Records and Accessibility of Common Data Base:** Sometimes due to hardware faults data records on direct access storage are destroyed. Such type of problem may be overcome by maintaining a duplicate record of the data on the back-up-tape. The original data records may be stored on the back-up tape as well as on the well as on the direct access storage device. For updating or affecting any change on the direct access storage, it is necessary to record such change on both in direct access storage and back-up-storage. The details of any change affected on the data are also recorded on the log tape. The log tape in fact contains the details, the

details about data, the data of change of data, person who affected the change, etc. In the event of spoilage of the direct access records, the back-up-storage records, along with the log tape may be utilized for creating a new file.

- (iii) **Organizing Data in Database:** To organize data collected from different sources/ data bases to evolve a data base for the organization, it is necessary to develop data standards. The development of data standards involves assigning of the data names, defining of data items, specifying the uses to which the data may be put and how it may be used by computer professionals and users. It also involves defining of the relationships between different data items and the way in which it may be physically stored and maintained.
- (iv) **Maintenance of Database Performance.** The performance of database is measured as the response time of the database or the time lapse between initiation of a request and the receipt of the reply from the common database. This type of measuring performance of the database is not appreciated because the response time of the database may be less but still the executives may take sufficient time to take decisions. Thus, now a modified version of measuring performance of database has gained popularity. Under modified versions, performance is measured in terms of the time jointly taken by the database to respond and the executive for taking decision.

TYPES OF DATA BASES:

The growth of distribution processing, end user computing, decision support and executive information systems has caused the development of several types of databases; some of these are discussed as under:

1. Management Databases:

These databases store data and information extracted from selected operational and external databases.

- (i) It is created to meet information requirements for middle-level management also known as tactical level. Help us to produce information to be used for decision – making process.
- (ii) Its objective is to help the managers to carry out various functions such as planning, organizing, staffing, directing and controlling.
- (iii) The middle-level managers need (a) internal information which can be created by logically relating different physical databases at the operational level (b) external information from the other organizations, e.g., pricing policy of a competitor or import/export policy of the government, etc.

2. Operational Databases:

These databases also known as subject area database (SADB), transaction databases and production databases store detailed data needed to support the operations of the entire organization. It deals with meeting the requirements of operational level of management.

- (i) The main objective is (a) collection of data, (b) storage of data, (c) retrieval of data, and (d) processing of data to generate information to be used during the design of MIS.
- (ii) Separate databases are created for each function within the organization, e.g.,
(a) Finance, (b) Inventory, (c) Production, (d) Marketing, etc.

3. End User Databases:

These databases consist of a variety of data files developed by end users at their workstations.

- (i) The end-users are senior executives, i.e., top level of management.
- (ii) For each user menus are created during the design of external implementation of a database.
- (iii) Each user can create an additional database for keeping such data which is exclusively required for his own function. It may contain (a) electronic mails received from others, (b) spreadsheet containing data arranged in the form of a table, and (c) word processing documents, e.g., various letters written by the user, etc.

4. Distributed Databases:

These are databases of local work groups and departments at regional offices, branch offices, manufacturing plants and other work site, which can include segments of both

common operational and common user databases as well as data generated and used only at a user's own site.

- (i) It is a technique by which databases can be divided into multiple files so that separate databases can be physically created at different branch offices but logically connected to a single database with the help of defining pointers (links).
- (ii) It facilitates accessing the data from any location, e.g., data of a branch office can be accessed from the head office or vice-versa.
- (iii) Based upon the receipt of a transaction, all relevant databases can be simultaneously updated.
- (iv) It is best achieved with the help of relational database structure.

5. External Databases:

Access to external, privately owned online databases or databanks is available, for a fee, to end users and organizations from commercial information services.

- (i) Any organization needs data from external sources.
- (ii) These database may be in the form of (a) technical database giving specifications of various materials and services which are being purchased by the organization, (b) commercial database giving pricing details and other terms and conditions, (c) abstracts taken from various newspapers and technical magazines, etc.
- (iii) Most of these databases are privately owned which can be accessed either free of charge or at prescribed rates.

6. Analytical Databases:

These databases, also called multi-dimensional databases, store data and information extracted from selected operations and external databases. They consist of summarized data and information most needed by a manager and other end users in the organisation and are accessed by the online analytical processing systems, decision support systems and execute information systems.

7. Text Databases:

Text databases are a natural outgrowth of the use of computers to create and store documents electronically. Thus, online databases services store bibliographic information such as publications in large text databases. Text databases are also available on CD-ROM optical disks for use with microcomputer systems. Major corporations and government agencies have developed large text databases containing documents of all kinds. They use text databases management system software to help, create, store, search, retrieve, modify, and assemble documents and other information stored as text data in such databases.

8. Image Databases:

A wide variety of images can also be stored electronically in image databases, e.g., electronic encyclopedias are available on CD-ROM disks which store thousands of photographs and many animated sequences as digitized images, along with thousands of pages of text. The main appeal of image databases for business users is in document image processing. Thousands of pages of business documents, such as customer correspondence, purchase orders and invoices, as well as sales catalogues and service manuals, can be optically scanned and stored as document images on single optical disk. Image database management software allows employees in many companies to quickly retrieve and display documents from image databases holding millions of pages of document images. Workers can view and modify documents at their workstations and electronically forward them to the workstations of other end users in the organization.

DATABASE USERS

- (i) Naive users who are not aware of the presence of the database system supporting the usage,
- (ii) Online users who may communicate with database either directly through online terminal or indirectly through user interface or application programs. Usually they acquire at least some skill and experience in communicating with the data base.
- (iii) Application programmers who are responsible for developing the application programs and user interfaces,

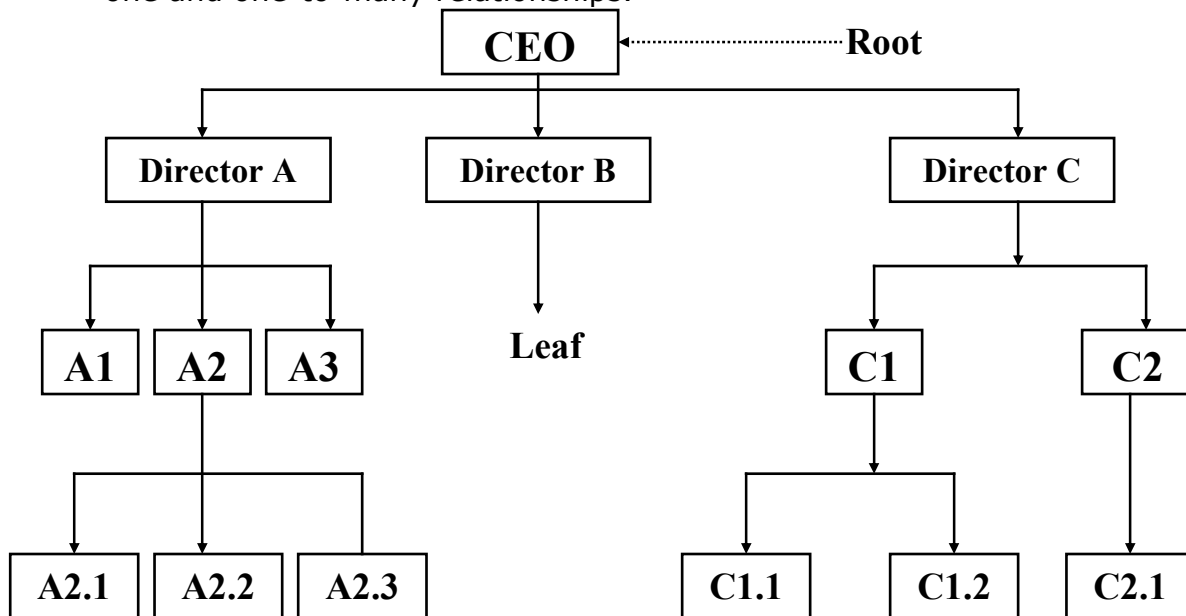
- (iv) Data Base Administrator who can execute centralized control and is responsible for maintaining the database. He is most familiar with the database.

DATA BASE STRUCTURES

Data is organised in the database in accordance with one of the three logical structuring techniques known as database structures. These models differ in the manner in which data elements or records are logically related and accessed.

A. HIERARCHICAL DATA BASE STRUCTURE:

Records are arranged according to the pattern generally followed in the constructions of an organisation chart i.e., inverted tree. Top record in the structure is known as root or master or parent record. Root may have several nodes or child records which constitute the next level of hierarchy. A child becomes a parent in respect of the next lower level of nodes in the hierarchy. When a parent record has no child, it is referred to as leaf. Each parent record may have one or more child record, but no child record may have more than one parent record. Thus, the hierarchical data structure implements one-to-one and one-to-many relationships.



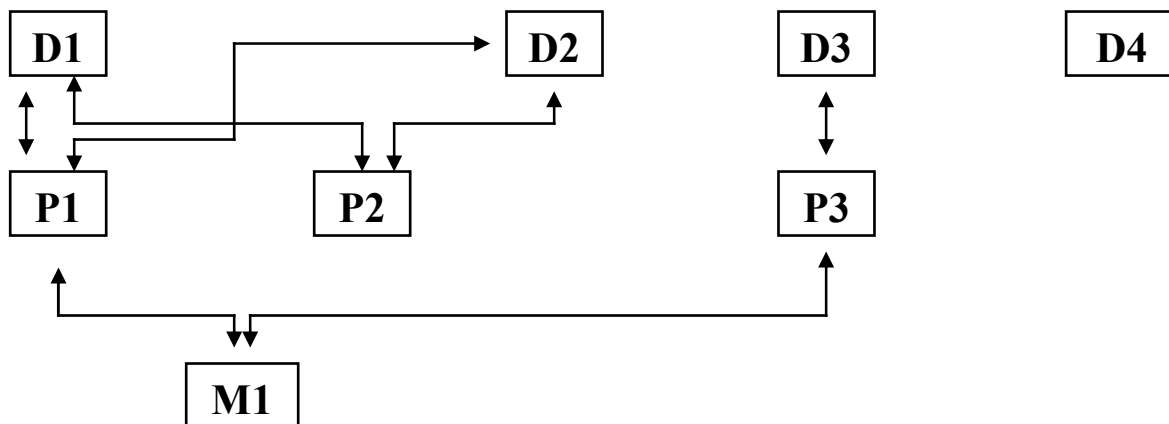
Features of Hierarchical data structure:

1. Simplicity as the parent-child concept is readily understood by most persons.
2. Hierarchical database management system usually processes structured, day to day operational data rapidly. In fact, the hierarchy of records is usually specifically organised to maximise the speed with which large batch operation such as payroll or sales invoice are processed.
3. If the parent node is deleted, then all the child nodes get deleted automatically. Conversely, a child node cannot be added without first adding the parent node.
4. For seeking information contained in any record the tracing process must start at the top with navigation continuing downwards until the requisite record is located.
5. Managerial use of query language to solve the problem may require multiple searches and prove very time consuming. Analysis and planning activities, which frequently involve ad hoc management queries of the database, are not effectively supported by a hierarchical DMBS.

B. NETWORK DATA STRUCTURE

The network data structure is also a kind of hierarchical organisation of records but with certain distinctive characteristics. In this case each node may have several parents. The nodes may be connected in a multidirectional manner. The connections are by means of pointers, i.e., records can be accessed in the forward as well as backward direction. These pointers are established when one of the fields in each record is used to contain address of the physical medium of the next record in the logical sequence. With appropriate DBMS software, it becomes possible to reach any node, for information retrieval, through more than one path. Once a record within a network has been retrieved, it also becomes possible to retrieve any number of records, depending on the relationship that is of interest to the user.

RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY



Features of Network Data Structure:

1. Information stored in a network data structure is often easier to restructure than in the case of hierarchical structure. This is because the path of data relationship is named in network data structure which make it possible to store member records in an arbitrary manner.
2. The complex relationship it creates becomes a storehouse of useful and easily accessible information.
3. Relevant for modelling many-to-many relationships

C. RELATIONAL DATA STRUCTURE

Relational data structure treats data as if they have been stored in two dimensional tables. It relates data stored in one table to data stored in another table so long as both the tables share a common data element.

In a relational model, each row of tables corresponds to a single record of the data base (a table row is called a tuple). The columns of a table break each tuple into distinct sections. These sections correspond to the field within a record. In relational terms, these fields are called attributes.

The relational data structure does not have a predetermined set of relationship between records. Instead, relational operators allow the user to define relationship at run time.

Features of Relational Data Structure:

1. High flexibility in programming and data retrieval.
2. Easier to use because it enables a computer system to accommodate a variety of inquiries in an efficient manner.
3. Processing time and storage space are relatively large in relational database.
4. Ease in data addition, and updation.
5. The links between data elements in a relational data base do not need to be made explicit at the time the database is created since new links can be structured at any time.
6. Because they do not specified the relationship among data elements in advance, relational data bases do not process large batch applications with the speed of other data bases.

DATABASE MANAGEMENT SYSTEMS

Definitions:

1. A database management system is simply the software that permits an organisation to centralize data, manage them efficiently, and provide access to the stored data by application programs.
2. A collection of programs required to store and retrieve data from a database is called a 'Data base Management System' (DBMS). It is a complex software system which constructs, expands and maintains the database and also provides the interface between the user and the data in the base.
3. DBMS is a software system which manages a database and protects it against damage and unauthorized access.

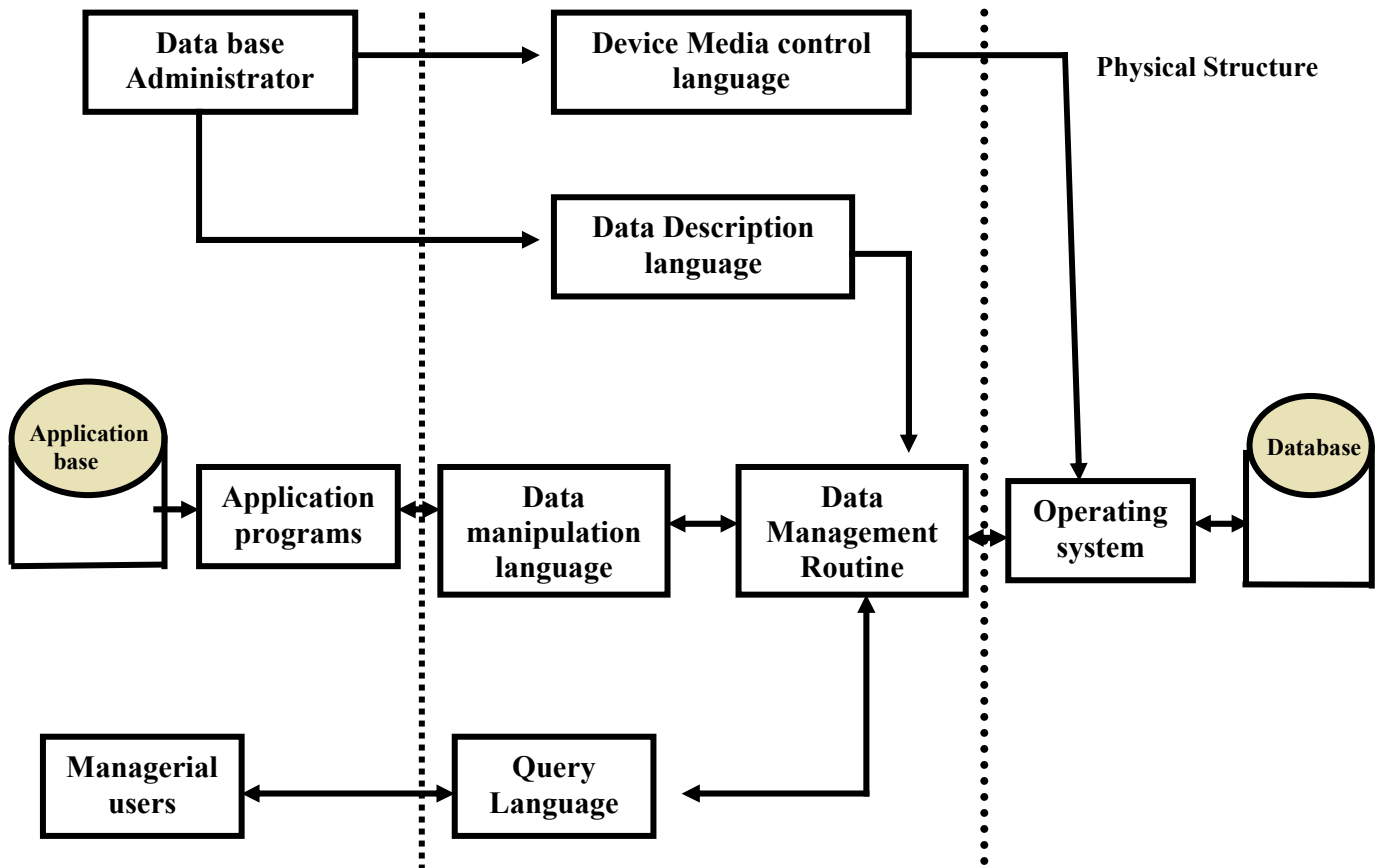
A database management system (DBMS) works like a computerized file cabinet. It stores and retrieves information that can be organized as a set of records. Each record

RELIABLE CLASSES/P.C.C./ INFORMATION TECHNOLOGY

may contain field of information. DBMS programs allow one to enter, edit, manipulate, view, modify, or play around with data contained in one or more files.

With a DBMS package, the following functions may be performed:

1. Enter, inspect and modify database records.
2. Protect data against unauthorized access.
3. Select records from the database on the basis of the data values in one or more fields.
4. Create the database.
5. Maintain the contents in an upto date manner.
6. Safeguard the data from loss or destruction.
7. Provide recovery and restart facilities after a hardware or software failure.
8. Provide interface with user programs written in different languages.
9. Print reports on the contents of the database.



Components of a database management system

CHARACTERISTICS OF DATABASE MANAGEMENT SYSTEMS:

Characteristics of a DBMS listed below, have important implications for its use:

1. The data are combined to form operational units to minimize the duplication of data and increase access to all data in the database.
2. The advancement in the database enables to add more data and program to the system.
3. The control in the system limits access to the database and builds the confidentiality of data in the database.
4. It has capacity to interrogating data files, retrieving and modifying data and recording the changes.
5. DBMS has the capacity to store large amount of data necessary for users' needs. They are stored on direct accessible devices.
6. It organizes data according to the specifications placed in the Data Definition language by the Database Administrator (DBA) or system designer in the initial design as well as when the Database Configuration changes; in the most suitable manner for each application.
7. It orchestrates and integrates data. As the basic element, the field level data is chained, interrelated or linked so that they can be juxtaposed in any desired

combination during execution of a desire application program. Thus DBMS is the vehicle to collect, combine and return a part of the available data to the user.

8. Serves as a filter between application programs and data associated with them. The application logic is divorced from the Input/Output logic.
9. Retrieves data in the required manner.
10. Data is secured and protected both in contents and relationships.
11. The operating system causes the data to be accessed, read, and transmitted to a buffer storage area in primary storage.

Database Software:

It enables people to access complex file systems with a minimum of difficulty. This is made possible by four components of database software.

1. Data Definition Language (DDL)
2. Data Manipulation Language (DML)
3. Query Language
4. Report Generator.

DATA DEFINITION LANGUAGE

Database scheme is specified by a set of definitions which are expressed by a special language called a Data Definition Language (DDL) which is also a link logical and physical views of the database.

Logical views refer to the way the user view the data, physical view refers to the way the data are physically stored. The logical structure of database is called a schema. A sub-schema is the way a particular application views the data from the database. Different program uses a set of DDL statements to construct a sub-schema that includes only data elements of interests.

The DDL is used to define the physical characteristics of each record, the fields within the record, and each field's logical name, data type, and length. The DDL is also used to specify relationships among the records.

The result of compilation of DDL statements is a set of tables which are stored in a special file called data dictionary (or directory). A data directory is a file that contains meta-data, i.e., data about data. This file is consulted before actual data is read or modified in the database system. The storage structure and access method used by the database system are specified by a set of definitions in a special type of DDL called a data storage and definition language.

A DDL is the basic tool for creating and maintaining a data dictionary. It is also used to regulate the rights of access (These rights are specified by the database administrator).

DATA MANIPULATION LANGUAGE

The data records files and logical structure in a shared database environment are processed by utilising another special programming language called Data Manipulation Language (DML) sometimes, referred as database command language.

It is a language that enables users to access or manipulate data.

A DML is not in itself a complete language. It is a set of statements that are embedded in host and query language.

DML is used by users to manipulate data in the database. It includes all the commands that enable the users to manipulate the data, and users can view the data, add new data, delete existing data, and modify selected fields in a record.

QUERY LANGUAGE

This language enables the users to ask specific questions to the database.

Database query languages are not universal. They are specific to the database management system with which they are used. Thus, each database may have its own query language with unique rules and instruction format.

The most popular query language is structured query language (SQL) which has an industry-wide standard. SQL is used in database management systems on many platforms.

REPORT GENERATOR

It enables the users to design and generate reports and graphics. The procedures for formatting a report and performing sub-totals and breaks for each group of items, page

breaks, page headings on first and subsequent pages, grand totals, etc. can be quite complex. Yet they follow fairly regular rules. The regularity of the rules and procedures for report layout is the basis for a report generator. Using a report generator, the programmer describes the format of the report and characteristics of the data. The detailed procedures are generated by the software.

DATABASE ADMINISTRATOR

The database administrator heads a central group which organises and controls the database. He is responsible for assuring the integrity and security of the database, managing all activities relating to schema (data definitions, etc.), controlling validity and access rights for sub schemas as well as for preparing the requisite documentation.

Responsibilities of database administrator

1. Establishes appropriate content and format of data records, the structure of data relationships, and the appropriate data names and key fields.
2. Controls and monitors data base usage through the assignment of user passwords and the collection of statistics on utilisation of data and programs.
3. Concerned with efficient use of physical data storage equipment.
4. Guiding the initial design of the database and further developing and extending the data base as and when it becomes necessary.
5. Demonstrating the benefits of the database technology to the other managers of the organisation and to educate them in the use of technology.
6. Establishing and implementing database standards.
7. Deciding on the contents of database.
8. Establishing and monitoring database control and security policies and procedures designed to ensure that database updating is controlled and accurate, data is available to users with a bonafide need and it is not available to unauthorise users, to ensure that lost or destroyed data can be recovered.
9. Ensures that the operating staff performs its database processing related responsibilities which include loading of the database, taking backups, scheduling the database for use, following maintenance and security procedures, etc
10. Monitors the database environments.
11. Responsible for incorporating any enhancements into the database environment which may include new utility programs or new systems releases, and changes into internal procedures for using data base, etc.
12. Maintaining manuals for users describing how to use the facilities of the database.
13. Maintains a dictionary describing the data items.
14. Acts as a file designer and database consultant to other users in the organisation.

DATA MINING

Data mining is a process of extracting hidden patterns from the data. As more data is gathered, with the amount of data doubling every three years, data mining is becoming an increasingly important tool to transform this data into information.

Data mining is the process of using computing power to convert captured data into information and knowledge. Data mining identifies trends within data so that users have the opportunity to identify key attributes of processes and target opportunities.

The term data mining is often used to apply to the two separate processes of knowledge discovery and prediction. Knowledge discovery provides explicit information about the characteristics of the collected data, using a number of techniques. Forecasting and predictive modelling provide predictions of future events.

PROCESS OF DATA MINING

Knowledge Discovery in Databases (KDD) is the name given to the process of using data mining algorithms.

Once the objective for the KDD process is known, a target data set must be assembled. The target set is then cleaned. Cleaning removes the observations with noise and missing data. The cleaned data is reduced into feature vectors. A feature vector is a summarised version of the raw data observation. The feature vectors are divided into two sets, the training set and the test set. The training set is used to train the data

mining algorithm(s), while the test set is used to verify the accuracy of any patterns found.

Data Mining commonly involves four classes of task.

- Classification – Arranges the data into predefined groups.
- Clustering: Is like classification but the groups are not predefined, so the algorithm will try to group similar items together.
- Regression- Attempts to find a function which models the data with the least error.
- Association rule learning: Searches for relationships between variables.
For example a supermarket might gather data of what each customer buys. Using association rule learning, the supermarket can work out what products are frequently bought together, which is useful for marketing purposes.

The final step of KDD is to evaluate the patterns produced by the data mining algorithms.

The evaluation uses a "test set" of data which the data mining algorithm was not trained on. The learnt patterns are applied to this "test set" and the resulting output is compared to the desired output.

If the learnt patterns do meet the desired standards then the final step is to interpret the learnt patterns and turn them into knowledge.

DATA WAREHOUSE

Data warehouse is a repository of an organisation's electronically stored data. Data warehouses are designed to facilitate reporting and analysis.

This definition of the data warehouse focuses on data storage. However, the means to retrieve and analyze data, to extract, transform and load data, and to manage the data dictionary are also considered essential components of a data warehousing system.

A data warehouse is a computer data base that collect integrates and stores data with the aim of producing accurate and timely management information and supporting data analysis. Data warehouses bring in data from a range of data sources such as mainframe computers mini computers as well as personal computer and office automation software and integrate this information in a single place. This alongwith user friendly reporting tools has led to the growth of this types of computer system.

Benefits of Data Warehousing:

1. Provides a common data models for all data of interest regardless of the data's source.
2. Makes it easier to report and analysis information than it would be it multiple data models for used to retrieve data.
3. Prior to loading data in the data warehouse. Inconsistencies are identified and resolved. This simplifies reporting and analysis.
4. Information in the data warehouse can be stored safely for extended periods of time.
5. They are separate from operational systems, so, data warehouses provide retrieval of data without slowing down operational systems.
6. Facilitates Decision Support System application such as trend reports, exception reports and reports comparing actual performance versus goals.

Disadvantages of Data warehouse:

1. Data warehouses can have high cost of maintenance.
2. Data warehouses can get outdated relatively quickly

BACKUP & RECOVERY

Backup is a utility program used to make a copy of the contents of database files and log files. The database files consist of a database root file, log file, mirror log file, and other database files called dbspaces.

Recovery is a sequence of tasks performed to restore a database to some point-in-time. Recovery is performed when either a hardware or media failure occurs. Hardware failure is a physical component failure in the machine, such as, a disk drive, controller card, or power supply. Media failure is the result of unexpected database error when processing data.

Certain issues related with Database Backup and Recovery

A Transaction Log is a file that records database modifications. Database modifications consist of inserts, updates, deletes, commits, rollbacks, and database schema changes. A Mirror Log is an optional file and has a file extension of .mlg. It is a copy of a transaction log and provides additional protection against the loss of data in the event the transaction log becomes unusable.

An Online Backup is performed by executing the command-line or from the 'Backup Database' utility. When an online backup process begins the database engine externalizes all cached data pages kept in memory to the database file(s) on disk. This process is called a checkpoint. The database engine continues recording activity in the transaction log file while the database is being backed up. The log file is backed up after the backup utility finishes backing up the database. The log file contains all of the transactions recorded since the last database backup. For this reason the log file from an online full backup must be 'applied' to the database during recovery.

An Offline Backup does not have to participate in recovery but it may be used in recovery if a prior database backup is used.

A Live Backup is carried out by using the BACKUP utility with the command-line option. A live backup provides a redundant copy of the transaction log for restart of your system on a secondary machine in the event the primary database server machine becomes unusable.

A Full Backup is the database backup utility copies the database and log.

An Incremental Backup uses the DBBACKUP utility to copy the transaction log file since the most recent full backup. When you perform an incremental backup the mirror log is not backed up.

Developing a Backup and Recovery Strategy

The steps suggested in the development of a backup and recovery strategy consist of the following:

- Understand what backup and recovery means to you business.
- Management commits time and resources for the project
- Develop, test, time, document, health check, deploy, and monitor
- Beware of any external factors that affect recovery
- Address secondary backup issues.