

4. Data Storage, Retrieval & DBMS

Q.No.1. Write about different coding schemes.

BCD:

- » BCD stands for binary Coded Decimal system.
- » This is a 6 bit code i.e. each character is expressed in terms of a 6 bit code.
- » There will be 64 ways of representing digits, alphabets and special symbols.
- » There is no parity bit in this coding scheme.

ASCII:

- » ASCII stands for American Standard Code for Information Interchange.
- » This code is extensively used in small computers, peripherals, instruments and communication devices.
- » It is a seven-bit code and has replaced many of the special codes that were previously used.
- » Microcomputers using 8-bit word length use 7 bits to represent the basic code. The 8th bit is used for parity or it may be permanently 1 or 0.
- » With 7 bits, up to 128 characters can be coded.
- » A letter, digit or special symbol is called a character. It includes upper and lower case alphabets, numbers, punctuation marks, special and control characters.
- » Here the first 3 bits are zone bits and the remaining 4 bits represent digit values.

ASCII-8:

- » A newer version of ASCII is ASCII-8 code, which is an 8-bit code.
- » With 8 bits, the code capacity is extended to 256 characters.

EBCDIC:

- » EBCDIC stands for Extended Binary Coded Decimal Interchange Code.
- » It is a standard character code for large computers.
- » It is a 8-bit code without parity and 9th bit can be used as parity bit.
- » Using 8 bits 256 characters can be coded.
- » In ASCII-8 and EBCDIC, the first 4 bits are known as zone bits and the remaining 4 bits represent digit values.

Character	ASCII	EBCDIC	F	01000110	11000110
0	00110000	11110000	G	01000111	11000111
1	00110001	11110001	H	01001000	11001000
2	00110010	11110010	I	01001001	11001001
3	00110011	11110011	J	01001010	11010001
4	00110100	11110100	K	01001011	11010010
5	00110101	11110101	L	01001100	11010011
6	00110110	11110110	M	01001101	11010100
7	00110111	11110111	N	01001110	11010101
8	00111000	11111000	O	01001111	11010110
9	00111001	11111001	P	01010000	11010111
A	01000001	11000001	Q	01010001	11011000
B	01000010	11000010	R	01010010	11011001
C	01000011	11000011	S	01010011	11100010
D	01000100	11000100	T	01010100	11100011
E	01000101	11000101	U	01010101	11100100
			V	01010110	11100101
			W	01010111	11100110
			X	01011000	11100111
			Y	01011001	11101000
			Z	01011010	11101001

Q.No.2. Define the term key. Write about different types of keys?

Key: A key is a set of one or more columns whose combined values are unique among all occurrences in a given table. A key is the relational means of specifying uniqueness.

Types of Keys:**a. Candidate Key:**

- » It is a set of one or more columns whose combined values are unique among all occurrences (i.e. tuples or rows)
- » Since a null value is not guaranteed to be unique, candidate key can't be null.
- » There can be any number of candidate keys in a table.

b. Primary Key:

- » It is a candidate key of the table which the database designer designates as "primary key".
- » Entire data in a database is sorted using this primary key.
- » The primary key may be selected for convenience, comprehension, performance or any other reasons.
- » It must have unique values.

c. Alternate Key:

- » It is a candidate key which is defined as primary key but not currently selected for the purpose is called alternate key.
- » It is a function of all candidate keys minus primary key.

d. Secondary Key:

- » Secondary keys can be defined for each table to optimize the data access.
- » A candidate key which is not selected as a primary key is known as Secondary Key.
- » Like the primary key, the secondary key can consist of multiple columns.
- » A table is first sorted on the basis of primary key and then on the basis of secondary key.
- » They help to prevent sequential scans over the table.

e. Referential Integrity:

- » It is a feature provided by Relational Database Management Systems (RDBMS's).
- » It prevents users or applications from entering inconsistent data.
- » It is a form of range control in which the value of a field must be similar to the value of some other field in another row of the same or different table.
- » Many RDBMS's have various referential integrity rules that are applied when a relationship is created between 2 tables.
- » This is very much useful in case of fields that have repeated values.

Q.No.3. Define the terms data and information. State the relationship between Character, Data item, Record, File and Database?

Data: Generally the word "data" is used in singular sense but actually it is the plural form of "datum". Data can be defined as any fact, observation, assumption or occurrence. Data are compiled to form reports, figures documents etc. Data can be in the form of numbers or alphabetical characters or special symbols (=, + %, Rs. etc.). Some important data definitions are given below:

- a. Digit:** A simple numeric character 0, 1, 2.....9 is called digit.
- b. Character:** A single alphabetic, numeric or special symbol that is used to represent data. For e.g. 0, 1, 2 9, A, B.....Z, a, b,.... z, +, - , %, etc.
- c. Data item (field):** A set of characters which are used together to represent a specific data element. For e.g. name item consists of alphabetic characters & amount item consists of numbers.
- d. Record:** It consists of a group of data items, related to an object of data processing. For e.g., a payroll record consists of data fields such as name, age, qualification, sex, wage rate etc. **(N00)**
- e. Data file:** It is a group of related records maintained in some pre-arranged order. **(N00)**
- f. Database:** A group of related data files is known as database.

Information: Information is nothing but organised or classified data and it has some additional value to the user. For example, a file containing transactions occurred during a year, in a company, is data for a managing director. If the same file is arranged, processed and financial statements are prepared then it becomes information. Such statements are very much useful for a manager.

Q.No.4. What is meant by Data Processing? Write down the steps involved in data processing?
Data Processing:

- » It is a process of restructuring or reordering of data by people or machines to increase their usefulness and value to the user.
- » Data processing can be done in several ways. In olden days it was performed manually with the help of simple tools such as paper, pencil and filing cabinets.
- » Afterwards data processing was done electro-mechanically with the help of unit record machines.
- » Now it is done electronically with the help of a computer. With the help of computers and peripheral devices, sorting, classifying, summarizing and other similar activities can be performed at a very high speed.

Steps in Data Processing:

1. **Preparation of source documents:** The first step is to collect relevant facts and figures and to record them on source documents. For e.g. in a population survey - name, address, age, sex, occupation, etc. are first recorded on survey sheets or some other documents.
2. **Recording / Input Data:** Once data is given as input to a data processing system. The method of giving input will depend upon input media and devices.
3. **Processing of data:** Here, user may classify or sort the given data. For example, in a population survey, we can classify people on the basis of occupation or on the basis of age. We may sort the data in alphabetical order or in any other order. In all data processing, systems following fundamental operations are performed:
 - a. **Classifying:** It involves grouping of like items ^(=similar items) or transactions. Data are generally classified according to alphabetic or numeric code.
 - b. **Sorting:** Sorting means arrangement of data or transactions in ascending or descending order. Sorting may be done on numeric or alphabetic data. For example, Indian States may be arranged in the descending order of population or on the basis of name.
 - c. **Calculating:** It includes adding, subtracting, multiplying or dividing data to produce useful results. For example – computation of interest on amount etc.
 - d. **Summarizing:** It involves consolidating of data, emphasising main points. Generally, it involves finding out totals, calculating percentages etc.
4. **Information Output (Reporting):** The summarised data becomes information and it is presented to the management in the form of reports. Reporting is the final step of all data processing activities. For e.g. financial statements. Information should be understandable easily and quickly.
5. **Data Storage:** In most of the cases, the results of processing are retained for future use. For example, it may be necessary to compare the latest figures with previous figures. That is why data processing system requires a considerable amount of secondary storage space.

Q.No.5. Write about different types of data access methods? (N99-5M)

There are 2 principal ways to access data – Direct Access and Sequential Access.

Sequential Access:

- » Sequential Access refers to the method in which records are accessed in a sequence i.e. in the order in which they are stored. If a particular record is required to be accessed, all previous records should be passed.
- » Under sequential access method, all the records are kept in sequence either in ascending order or in descending order on the basis of key field.
- » For E.g.: Magnetic tape devices.

Direct Access:

- » In Direct access method (or Random Access Method), access of data is independent of the storage sequence. Individual records can be accessed at random, in no particular order or sequence.
- » Under this method, any record can be accessed at random without passing through intervening records.
- » For E.g.: Magnetic disk, Optical disk, etc.

Q.No.6. Write about different types of data/file organisation methods?**Serial File Organisation:**

- » This is the simplest file organization scheme. In this method, records can be arranged one after another in chronological order.
- » Serial organization is commonly found in transaction files.
- » Sometimes, records in a serially organized file are processed in the order in which they occur.
- » In this scheme, a new customer is given the next highest account number that has not been issued. Data about new customer are placed at the end of the existing customer account file.
- » When this approach is used, it is easy to distinguish old customers from new ones because old customers have lower account numbers.

Sequential File Organisation:

- » In sequential file organisation, all records are arranged in ascending or descending order by a record key. For example, in Payroll master file, records will be arranged in ascending order by employee number.
- » To locate an individual record, a sequential search is performed, beginning from the first record. Thus each individual record must be examined until the required record is located.
- » Such a search is time consuming. Therefore, sequential file organization can't provide immediate access to individual records.
- » Sequential organization is suitable for master files in a normal batch processing environment where all or most of the records are processed.
- » Certain storage media, like magnetic tape and paper tape allow sequential file organization only. However, direct access storage devices like floppy disk, CD, DVD etc. allow sequential file organization as well as direct access file organisation.

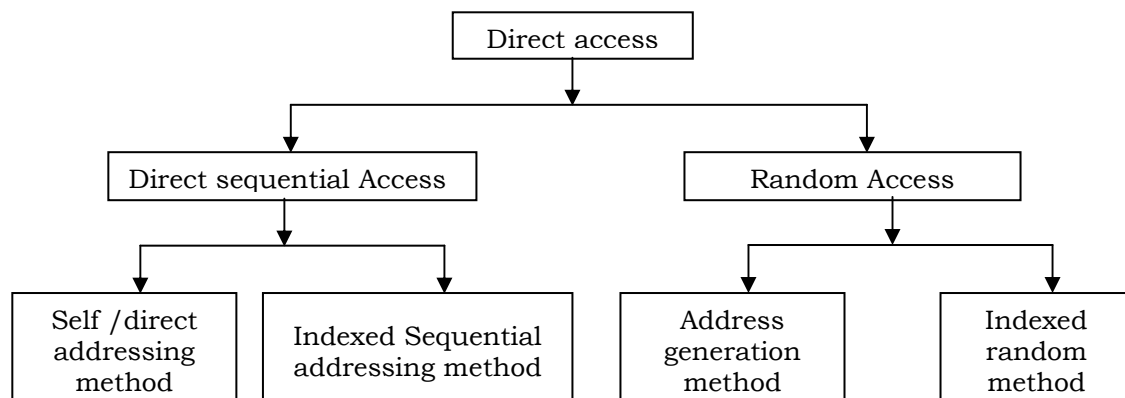
Advantages:

- » Easy to organise, maintain and understand,
- » For locating a particular record it is sufficient to specify the key field,
- » Relatively inexpensive media and devices can be used for storage and processing of files.
- » It is very efficient when the activity ratio is very high i.e. for an application where large number of records in a file are to be regularly updated at scheduled intervals. For e.g. applications such as payroll processing.

Disadvantages:

- » Inefficient and uneconomical for applications in which the activity ratio is very low.
- » It is required to read the entire sequential file even to update few records in the file.
- » Before processing, transactions must be stored and arranged in a sequence.
- » A batch of transactions must be accumulated before processing begins.
- » Very soon data becomes obsolete because processing starts only after accumulating a batch of transactions.

Direct Access File Organisation method: Direct access file organization allows immediate access to individual records on the file. There are some devices which can provide direct access and are called Direct Access Storage Devices (DASD). They can directly reach any required location. Discs are the most widely used direct access storage devices. Most widely used direct access techniques are shown below:



Direct Sequential Access Methods:**A. Self (Direct) Addressing method:**

- » Under self (direct) addressing system, a record key is used as its relative address.
- » Therefore, we can compute the record's address directly from the record key and the physical address of the first record in the file.
- » This method is suitable for fixed length records.
- » Suppose we want to store 1,60,000 payroll records in a hard disc with 200 cylinders. The first cylinder carries the first 800 records, 2nd cylinder carries the next 800 records and so on. For processing the file, the read/write heads would move cylinder by cylinder in which records are sequentially arranged. For example, ten faces in the first cylinder would carry the first 800 records as below:

F _{1,1}	1 to 80
F _{1,2}	81 to 160
	:
	:
F _{1,10}	721 to 800

Advantage: There is no need to store an index because the data itself is stored in a particular sequence.

Disadvantages:

- » This method is highly impractical because files have gaps in the keys and this would leave too many empty sectors.
- » The records must be of fixed length i.e. not suitable for variable length records.
- » If some records are deleted, their storage space remains empty.

B. Indexed Sequential File Organisation: (M96 – 5M, RTP)

- » The indexed sequential file organization or Indexed Sequential Access Method (ISAM) is a hybrid between sequential and direct access file organisations.
- » The records within the file are stored sequentially but direct access to individual records is possible through an index. This index is similar to a card catalog in a library.
- » This organization permits both sequential as well as random access.
- » For direct access of a record, the key is located by the operating system, in the index file and the address so obtained is used accessing the record. For sequential access, the index is not required.
- » If it is needed to add additional records, they can be stored in a special overflow area.
- » It combines the best features of sequential file organization and direct access.

Advantage:

- » In this method it is not required to arrange the transactions in a sequence because an index is used to locate the address.
- » If the activity ratio is high it is economical to use sequential processing of records.
- » If the activity ratio is low it efficient to use direct access processing of records.

Disadvantages:

- » Relatively expensive.
- » Requires extra storage space to make subsequent insertions.
- » Requires extra storage space for storing indexes.
- » Lot of time is wasted in processing the indexes and in reordering the files periodically.

Random Access Organisation: (N07-5M) In this method, transactions can be processed in any order and written at any location throughout the file. To access a record, prior records need not be examined first. Generally this method is used by CPU. The CPU can go directly to the desired location using randomising procedure.

Under randomising procedure records are stored in such a way that there does not exist any relationship between keys of adjacent records. This technique provides for converting the record key number into physical location represented by a disk address through computational procedures.

Q.No.7. Describe various factors which determine the best file organization (Or) With reference to computers explain briefly the best file organization. (M01-3M, M03, N06, M08 – 5M, RTP)

Several factors must be considered in determining the best file organization. These factors are:

- File volatility:** It refers to number of additions and deletions to the file in a given period of time. For example payroll file of a construction company, where employees are constantly changing, is a highly volatile file. In such a situation ISAM is not a good choice and other direct access methods would be better. If there are no interrogation requirements, even sequential file organisation would be appropriate. (M08-1M)
- File activity:** It is the proportion of master file records that are actually used or accessed in a given period of time. Some times each transaction will be processed immediately and hence, at a time, only one master record is accessed. In such a situation direct access method is best suited. At some other times almost every record in master file is accessed. For example, when weekly payroll is processed, every record in payroll master file is accessed. In such a situation, it is sufficient to arrange the file sequentially.
- File interrogation:** It refers to retrieval of information from a file. If the retrieval of individual records must be fast, to support a real-time operation (such as airline reservation) then some kind of direct access file organization is required. On the other hand, if requirements for data can be delayed then all the individual requests can be processed under batch processing mode. In such a situation, sequential file organisation is best suited.
- File Size:** Large files that require many individual references to records, with immediate responses must be organised under some type of direct access method. In case of small files it may be more efficient to search the entire file sequentially.

Q.No.8. Explain the short comings of traditional file processing system. (Or) List out the shortcomings of conventional file oriented data processing system that have been overcome by database management system. (Or) Management problems of file processing system. (M05, M07-5M, RTP)

Traditional information systems were designed to meet specific information and data processing needs of a particular department such as accounting, sales, purchasing etc. Different files are created to support these functions. Data needed for each user application was stored in independent files. Many of the fields in these files were common. Following are some of the problems of traditional file processing system:

- Data redundancy:** Independent data files used to include lot of duplicate data i.e. same data is recorded and stored in several files. This creates lot of inconsistency among data stored in separate files. This data redundancy causes several problems when data has to be updated.
- Lack of Data Integration:** In file processing system, data produced by each application program is stored in a separate file. So, it is not possible to get integrated data (i.e. data generated from more than one program). Such kind of data is generally requested by managers for their decision making purposes.
- Data Dependence:** In file processing system, major components of a system i.e. organisation of files, their physical location in storage devices, hardware, application software, etc. depend on one another. If changes are made in the format or structure of data or records in a file then changes have to be made in all programs that use such file.
- Other Problems:** In file processing system, data elements such as stock numbers and customer addresses are defined differently by different end users and applications. It is very difficult to develop programs which access such data. In addition, the integrity (i.e. the accuracy and completeness) of data is suspected because there is no control over their use and maintenance.

Q.No.9. Briefly describe Database concepts (Or) Write about database processing / database? (RTP)

Introduction: The concepts of database and database management system were developed to solve the problems of file processing systems. A database is an integrated collection of logically related records and file. It consolidates records stored in independent files so that it serves as a common pool of data, to be accessed by different application programs. The data stored in a database is independent of computer programs using it and of secondary storage devices.

Definition: A database is a computer file system that uses a particular file organisation to facilitate rapid updating of individual records, simultaneous updating of related records, easy access to all records by all application programs and rapid access to all stored data which must be brought together for a particular routine report or inquiry or a special purpose report or enquiry. This is the precise definition of database given by "G.M.Scott". Each of the italicized phrases in the definition has a special meaning which helps to define a database.

- a. **Uses particular file organisation:** In a database, records are organized using any of the file organisation techniques available. It helps to establish associations in a database.
- b. **Rapidly updates individual records and simultaneously updates related records:** The database permits the entry of an individual transaction to update all records affected by that transaction simultaneously.
- c. **Easy access to all records by all application programs:** DBMS uses standard data definitions and record formats. With this one application program can access data from another application program. Without a database, each application program can access data from its own file only.
- d. **Rapid access to all stored data:** It provides 'rapid access' to all stored data needed by a "routine report or inquiry". Thus routine reports can be quickly provided after the end of the accounting period or whenever requested during the period because the processing of transactions is kept up to date.
- e. **Special purpose report or enquiry:** In DBMS records are kept continuously up to date. The structure of database files facilitate rapid development of special purpose reports about unanticipated problems.

Q.No.10. Write about the architecture of a database? (RTP)

Architecture of a Database: A database follows a three level architecture

- » External or user view,
- » Conceptual or global view,
- » Physical or internal view.

External or user view:

- » It is viewed by the chairman or the operation manager or the data entry operator,
- » It is at the highest level of database abstraction,
- » It includes only those portion of database or application programs which is of concern to the users,
- » It is described by means of a scheme, called the external schema,
- » It is defined by the users or written by the programmers.

E.g.: An external view in its Logical Record 1 may indicate employee name, employee address and in its Logical Record 2 may indicate employee name, employee address, employee code and employee salary.

Global or conceptual view:

- » It is viewed by Data Base Administrator,
- » All database entities and relationships among them are included,
- » Single view represents the entire database,
- » It is defined by the conceptual schema,
- » It describes all records, relationships and constraints or boundaries,
- » It includes data description to render it independent of the physical representation.

E.g.: A conceptual view may define employee code as a string of characters, employee address as a string, employee code as a key and employee salary as an integer.

Physical or internal view:

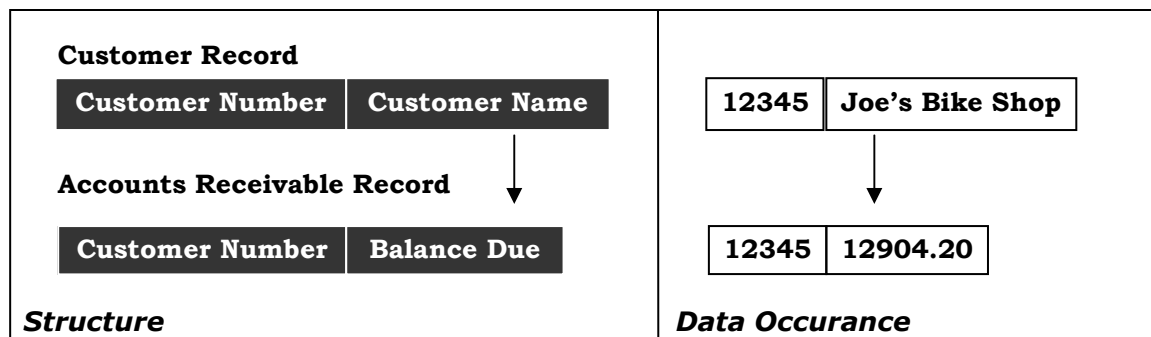
- » It is at the lowest level of database abstraction,
- » It is closest to the physical storage method,
- » It indicates how data will be stored,
- » It describes data structure,
- » It describes access methods,
- » It is expressed by internal schema.

E.g.: The internal view may define that employee name consists of 30 characters, employee address also consist of 30 characters, employee code consists of 5 characters and employee salary consists of 12 numbers.

Q.No.11. State different kinds of record relationships in a database?

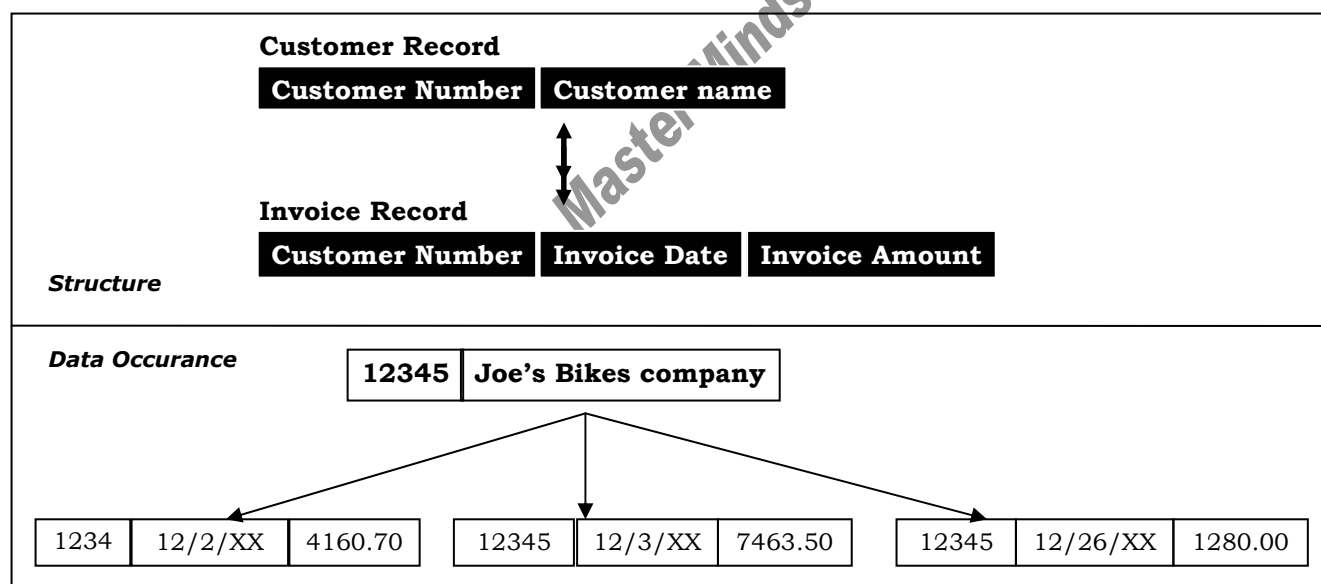
Record Relationships: The relationship between data is very important for a database to work. Related data are processed in the database through relationships. There are four types of relationships in databases.

One - One Relationship: In a one-to-one relationship, one group of data is related to only one other group of data and vice versa. For example, there are 2 files in a database – Customer Record file and Customer Receivables file. Customer Record file consists of following details – Customer number and customer name. On other hand Accounts Receivable file consists of following details - customer's number and balance due by the customer. There is one customer record for each customer and one accounts receivable record for each customer. Then, each customer record is related to one receivables record, and vice versa. Thus, there is a one-to-one relationship between customer records and accounts receivable records.



The above diagram shows a one-to-one relationship between two records. The first part of the figure shows the structure of the records. The arrow indicates one-to-one relationship between the records. The second Part of the figure shows data for the two records.

One to Many Relationship: The second type of relationship is known as one-to-many relationship. In this type, one group of data is related to one or more other groups of data but not vice versa. For example, assume that a database contains customer records and invoice records, each with an invoice date and amount. Each customer can have any number of invoices, but each invoice can be associated with only one customer. Then the relationship between customer records and invoice records is one-to-many.

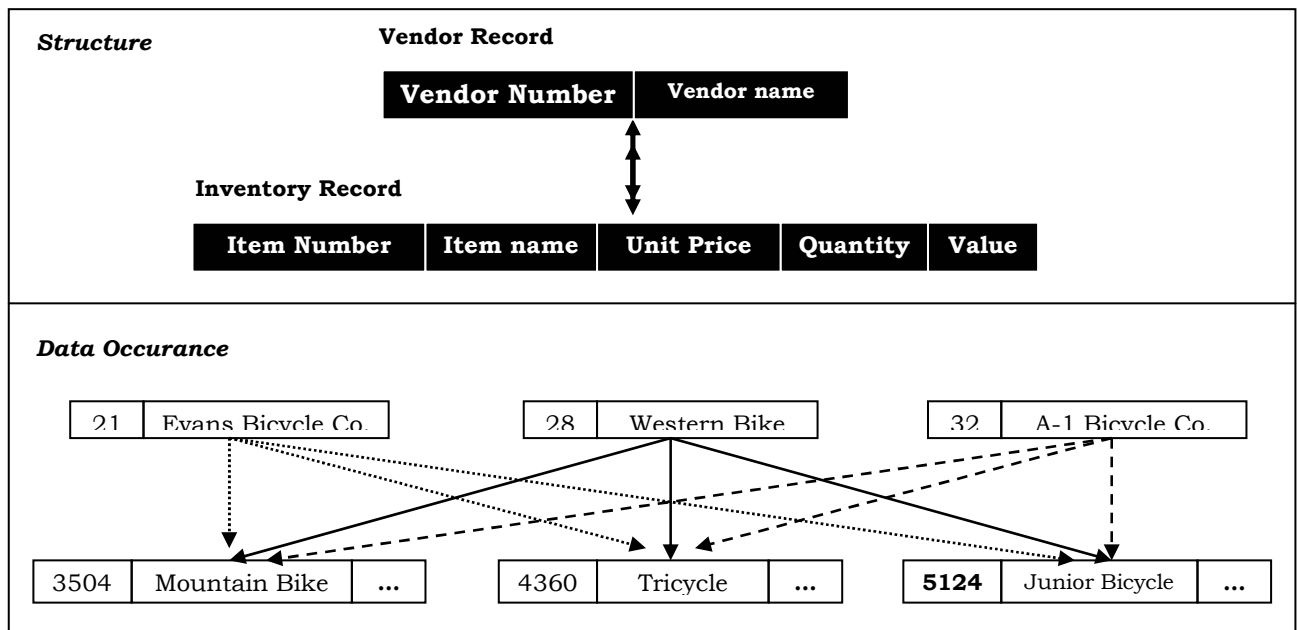


The above diagram shows the structure of customer records and invoice records and the one-to-many relationship between them. The arrow indicates one-to-many relationship. In this example, customer number 12345 has three Invoices.

Many to one Relationship: This is just reverse to One-to-many relationship.

Many-to-Many relationship: The last type of relationship is many-to-many relationship. In this relationship one or more groups of data are related to one or more other groups and vice versa. For example, assume that a database contains vendor records (each with vendor's name) and inventory records (each with item description, unit price and quantity on hand). Each vendor can supply several inventory items and each item can be supplied by several vendors. Then the relationship between vendor records and inventory records is many-many.

The figure shows the structure of the inventory and vendor records and the many-to-many relationship between them. The arrow indicates a many-to-many relationship. The below figure gives an example of these records and the relationship between them.



Q.No.12. What is meant by Database Structure? Write about different database structures available?

Database structure is the definition of data records in which information is stored, including the number of data fields. The data in a database is organized in such a way that it is easy for people to understand. The database organization is based on one of 3 approaches for arranging data in a database. The three approaches result in three types of databases:

- » Hierarchical databases,
- » Network databases and
- » Relational databases.

Q.No.13. Write about Hierarchical database structure? (M05, N07 – 5M)

Meaning:

- » In a hierarchical database, all relationships between groups of data are one-to-one or one-to many but not "many-to-many".
- » In this database structure, records are logically organised into a hierarchy of relationships and involve an inverted tree like structure.
- » The tree consists of hierarchy of nodes and the uppermost tree is called parent.
- » Every element can have any number of lower level elements, called children, but every node will have only one parent.

Features:

- » The hierarchy should be pre-determined and implemented and therefore, they are fixed in structure and are less flexible than other database structures.
- » Adhoc queries can't be made by the managers.
- » When the parent node is deleted, all the child nodes get automatically deleted.
- » If a particular record has to be traced then tracing will start from the root, continues downwards until the requisite record is located.

Advantages:

- » It facilitates day to day structured operations because it maximises the speed in case of batch operations.
- » Data processing becomes more efficient.
- » For most of the users it is simple to understand and can satisfy their information requirements.

Disadvantages:

- » It can't support many-to-many relationships.

- » Use of query languages require multiple searches and hence time consuming.
- » It is a rigid database structure and new relationships can't be accommodated within the structure.

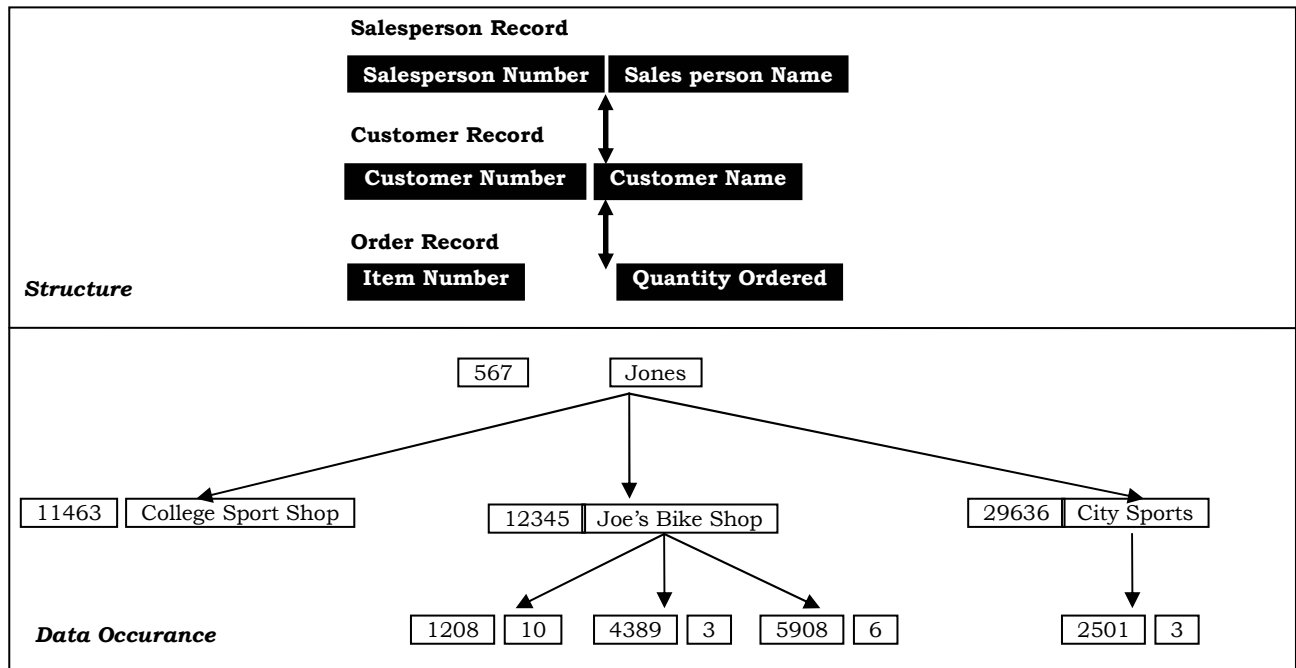


Illustration: Only one-to-many relationships are used in this database and all relationships go in the same direction. (One-to-one relationship can also be used but many-to-many relationships can't be used). The meaning of this database is that each salesperson can have any number of customers and that each customer can place any number of orders. The database in the figure looks like a family tree. In fact, this type of database is often called a tree. Going "down" the tree, each record has any number of "children," (i.e. related records). However, going "up" the tree, each record has only one "parent". Thus in the figure each salesperson can serve any number of customers but a customer can be served by one salesperson only.

Conclusion: DBMS's for hierarchical databases were first developed in 1960's. This type of databases is very good if data is naturally organised in a hierarchical manner, like a family tree. For example, data about products and parts used in the assembly of products is organized hierarchically: Each product is made up of several main parts; each main part is made up of several subparts and so on. A hierarchical database is appropriate for storing such type of data. However, hierarchical databases are not good if data is not organized in a hierarchical manner. For example, when many-to-many relationships are needed.

Q.No.14. Write about Network database structure?

In a **network database**, all types of relationships are allowed, without any restriction. Thus one-to-one, one-to-many and many-to-many relationships are permitted.

The below figure shows an example of network database. This database has one many-to-many relationship, which indicates that each vendor can supply many inventory items and each item can be supplied by many vendors. The database also has one one-to-many relationship which indicates that each warehouse can store any number of inventory items but each item can be stored in one warehouse only.

Features:

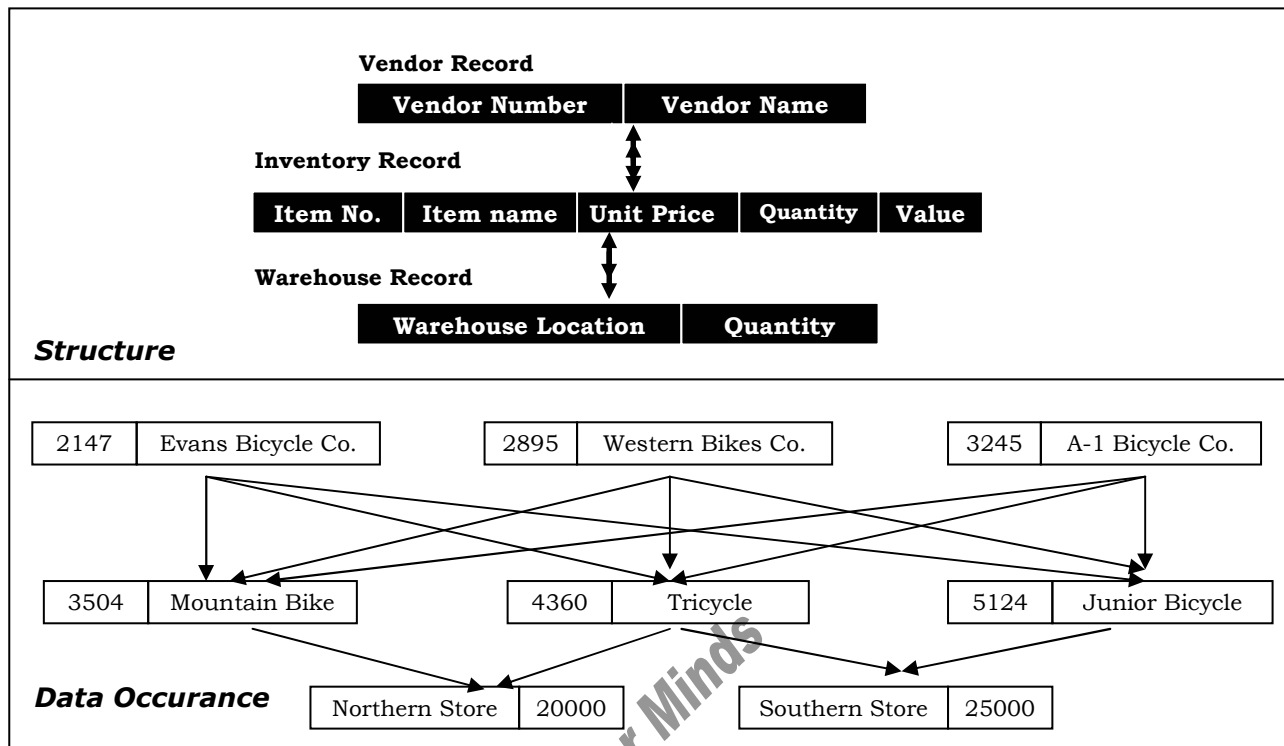
- » It is a modified version of hierarchial database. In network database structure each node may have several parents.
- » It is very difficult to develop this type of database structures.
- » It is useful for one-to-one and one-to-many record relationships.
- » The relationships should be pre-determined.

Advantages:

- » It is more flexible than hierarchial database structure.
- » They are appropriate to support data with many-to-many relationships.
- » Relatively efficient than hierarchial database structure.
- » Easy to restructure the information.

Disadvantages:

- » As the number of points in the network increases it is very difficult to operate and maintain this structure.
- » Less user friendly than hierarchical structure.
- » Very difficult to maintain because of its complex nature.
- » It is very difficult to add relationships to the existing database.



Conclusion: DBMS's for network databases were developed in 1970's. Network database is good for complex data that is not organised in a hierarchical way. For example, data about students, courses and faculty is complex: each student takes many courses, each course is taken by many students, each faculty member teaches many courses and each course can be taught by many faculty members. A network database would be appropriate for storing such type of data.

Q.No.15. Write about Relational Database Structure? (M07 – 5M)

A **relational database** organizes data in a different form than that of hierarchical and network databases. In a relational database, data is arranged in tables that have rows and columns. In a relational database, a table is called relation (not to be confused with 'relationship'), a row is called a **tuple** (which rhymes with couple) and a column is called an **attribute**. Some relational DBMS's use these terms; others use the terms table, row and column. Still others use file (for relation), record (for tuple) and field (for attribute).

Relationships in relational databases are different from that of relationships in hierarchical and network databases. A relationship exists between two relations (i.e. tables) because there is a common attribute in the relations. For example, there is a relationship between INVENTORY relation and ORDER relation because the ITEM NUMBER attribute is found in both relations. Similarly, there is a relationship between BILLING relation and ORDER relation because CUST_NUMBER attribute appears in both relations. Any type of relationship i.e. one-to-one, one-to-many and many-to-many can be represented in a relational database.

Advantages:

- » Highly flexible to program and retrieve data.
- » It is much easier to use.
- » Can handle queries in a more efficient way.

Disadvantages:

- » Processing efficiency is comparatively low.
- » Requires more processing capacity and memory.
- » Storage space requirements are high.
- » Processing can't be done without establishing the relationships.

Conclusion: The relational approach was first proposed in 1970, but DBMS's for relational databases become widespread until 1980. This type of database is very easy to understand because all data is presented in relations, (tables). Processing data in relations is relatively simple. If the database is large and complex, processing can take more time using the relational approach than with the hierarchical and network approaches.

Q.No.16. State different types of database users?

- a. **Naive users:** They are not aware of the presence of the database system supporting the usage,
- b. **Online users:** They communicate with database either directly through online terminal or indirectly through user interface or application programs. Usually they acquire atleast some skill and experience to communicate with the database,
- c. **Application programmers:** They are responsible for developing application programs and user interfaces.
- d. **Data Base Administrator:** He can execute centralized control and is responsible for maintaining the database. He is the most familiar person with the database.

Q.No.17. Why Data Base Management is being preferred in companies to Traditional File Processing system? (N02) (Or) Explain the advantages of DBMS? (Or) For what reasons do organisations chose to invest in database management systems? (RTP)

Companies find number of advantages in using database management system instead using file management systems. Following are important among them:

- a. **Data Sharing:** It is very easy to share between users. Data from the entire company can be used by users who need them. Managers can analyze the data in a more extensive ^(=wide) manner than in file management system.
- b. **Reduced data redundancy:** A database minimizes duplication of data from file to file. For example in a university database student's name and address may appear in only one record rather than in the files of many departments.
- c. **Data independence (M04-1M):** A database system keeps data separate from the application that uses the data. Thus, changes can be made to data definitions without making changes to every application program that uses the data. Thus program maintenance costs (i.e. cost of upgrading application programs) can be substantially reduced.
- d. **Increased user productivity:** In database environment user with minimum technical knowledge can also generate queries using query languages and can generate reports using report generators. To generate same queries/reports MIS departments take long periods of time. Thus productivity of user also increases.
- e. **Increased application programmer productivity:** Most database management systems offer application program development tools. With the help of these tools a programmers can write program code. These tools can be very powerful and can increase the productivity of programmer substantially.
- f. **Improved Data Integrity:** As data redundancy is minimized, data inconsistency and threat of non integrated data are substantially decreased.
- g. **Improved data administration & control:** In database environment the responsibility of protecting the database is placed in the hands of one person or department. It becomes easy to control access to data, privacy of data, updates, deletions of data etc.

Q.No.18. Discuss the 2 types of DBMS languages available in DBMS? (M07 – 5M, RTP)

Introduction: DBMS consists of 2 languages, Data Definition Language (DDL) and Data Manipulation Language (DML).

Data Definition Language (DDL): (N96 – 1M) DDL is a link between the logical and physical structure of the database. Here, logical means the way user views data; physical means the way data is physically stored. DDL is used to define the physical characteristics of each record i.e. fields within the record, each field's data type and each field's length and logical name. DDL is also used to specify relationships among records. The primary functions of DDL are:

- a. Describing the schema and sub schema.
- a. Describing the fields in each record.
- b. Describing the data type and name of each field.
- c. Providing data security restrictions.

- d. Providing logical and physical data dependence and
- e. Providing means for associating related records or fields.

Data Manipulation Language: Data Manipulation Language (DML) provides techniques for processing the database such as retrieval, sorting, display, and deletion of records. In summary, the functions of DML are:

- a. To provide techniques for data manipulation such as - deletion, replacement, retrieval, sorting or insertion of data or records.
- b. Enable the user and application programs to process data on logical basis rather than on physical location basis.
- c. Provide for independence of programming languages. DML should support several, high level procedural languages such as COBOL, PL/ 1, C ++ etc.
- d. Allow the user and application programs to be independent of physical data structures and database structure maintenance.
- e. It also helps to use the record relationships.

Q.No.19. Write about the structure of the DBMS?

- a. **DDL Compiler:**
 - » It converts data definition statements into a set of tables,
 - » Tables contain meta data (data about data) relating to database,
 - » It creates a format which can be used by other components of database.
- b. **Data Manager:**
 - » It is the central software component,
 - » It is also known as the database control system,
 - » It converts operations in users' queries to physical file system.
- c. **File Manager:**
 - » It is responsible for file structure,
 - » It is responsible for managing the space,
 - » It is responsible for locating blocks containing required records,
 - » It is responsible for requesting blocks from disk manager,
 - » It is responsible for transmitting required records to data manager.
- d. **Disk Manager:**
 - » It is a part of the Operating System,
 - » It carries out all physical input / output operations,
 - » It transfers blocks / pages requested by file manager.
- e. **Query Manager:**
 - » It interprets user's online queries,
 - » It converts them into an efficient series of operations,
 - » In a form it is capable of being sent to data manager,
 - » It uses data dictionary to find structure of relevant portion of database,
 - » It uses information to modify query,
 - » It prepares an optimal plan to access database for efficient data retrieval.
- f. **Data Dictionary:**
 - » It maintains information pertaining to structure and usage of data and meta data,
 - » It is consulted by database users to learn what each piece of data and various synonyms of data field means.

Q.No.20. Who is a Database Administrator? What are his functions? (Or) List out the functions of database administrator? (N96, M02, N03, M06 – 5M, RTP)

DBA is a database professional who actually creates and maintains the database and also carries out the policies developed by DA. DBA is responsible for defining the internal layout of the database and also for ensuring that the internal layout optimises system performance.

To design the database, DBA must know the data requirements of users. Then he can decide the schedule, accuracy requirements, frequency of data access, search strategies, physical storage requirements of data, level of security needed, response time requirements, etc. He may also identify the source of data and the

person responsible for originating and updating data. The database administrator then converts these requirements into a physical design. Following are other functions of a DBA.

- » To assist in the creation and maintenance of database.
- » To carrying out the policies developed by Data Administrator.
- » To look after the technical aspects of database management.
- » Defining internal layout of the database.
- » Designing a system as per the requirement of the organization.
- » Defining the contents of the database.
- » Using DDL to define the various data elements and their usage.
- » Maintain standards and control access to database are important aspects to be looked after by DBA.
- » To ensure proper preparation of documents to help the end user to use the database effectively.
- » To perform routine jobs like taking a back up, maintenance and security procedures, scheduling database following a restart and recovery procedure are done in a proper way.
- » To ensure conduct of periodic audit to check for identifying and correcting violation of standards, documents and correct errors.
- » To incorporate new system to access and control the database.
- » To manage all activities relating to the schema controls the validity and access rights for sub-schemas, and preparation of requisite documentation.
- » To ensure availability of data for sharing by the entire organisation and to ensure that the database is developed for management purposes.
- » To co-ordinate discussions between user groups to determine the contents and format of the database so that data redundancy is kept to minimum.
- » To demonstrate the benefits and the database technology to other managers of the organization and to educate them in the uses of DBMS.

Q.No.21. Write about different components required under a database environment? (N04 – 10M)

Database environment consists of following components:

- a. **The Database files:** These files have data elements stored in database file organisation formats. The database is created in such a way so as to achieve the objectives of speed, multiple access paths, minimum storage, program data independence, etc.
- b. **Database Management System (DBMS):** DBMS is a set of software programs that manages the database files. DBMS accesses the files, updates the records and retrieves the requested data. The DBMS is responsible for data security which is very important in a database environment because database will be accessed by number of users. **(M07 – 1M)**
- c. **The users:** Users consist of both traditional users (such as accounting personnel) and application programmers who are not traditionally considered as users. Users can directly interact with the database through query languages or indirectly through application programs.
- d. **A host language interface systems:** This is a part of DBMS which communicates with the application programs. It interprets the instructions in high-level language of application programs and tells to the DBMS.
- e. **The application programs:** These programs perform the same functions that they do in a traditional system but they are independent of data files and use standard data definitions. This independence and standardization makes the development process easier and faster.
- f. **A natural language interface systems:** With the help of this language, users can make queries to the database. This language is 'English-like' because instructions of this language are usually in the form of simple commands in English, rather than symbols and equations.
- g. **The data dictionary / directory:** Data dictionary is a centralized depository ^(=collection) of information, in a computerized form, about the data in the database. Data dictionary consists of schema of the database i.e. the name of each item in the database, a description and definition of its attributes along with the names of the programs that use them, and who is responsible for the data etc. These descriptions and definitions are called as "data standards". **(M07 – 1M)**

- h. **Online access and update terminals:** These may be adjacent ^(=near to) to the computer or even thousands of miles away. They may be dumb terminals, smart terminals or microcomputers.
- i. **The output system or report generator:** This provides routine job reports, documents and special reports.

Q.No.22. Write short notes on Data Dictionary / Data Directory. (M03, M05 – 5M)

- a. Data dictionary is a centralized depository of information, in a computerized form, about the data in the database.
- b. It is also called Data Directory and contains the schema of the database i.e. the name of each item in the database and a description of its attributes names of the programs that use them, who is responsible for the data; authorization tables that specify users and the data and programs authorized for their use. These descriptions and definitions are covered under data dictionary.
- c. The data dictionary is helpful in the following ways:
 - » If a data field is to be altered in any way or eliminated from the database, the dictionary can identify all programs that might require some modifications.
 - » It also helps to protect the integrity of data in the database.
 - » It helps programmer to quickly locate blocks of code that can be reused in new applications.
 - » It also ensures that all design or coding actions performed by an analyst are consistent with the application.
- d. They inform the user / programmer of what data is available in the database and how it is used.

Q.No.23. Write about different types of databases? (M06 – 10M)

- a. **Operational databases:** These databases contain detailed data needed to support the operations of the entire organisation. They are also called subject area databases or transaction databases or production databases. Examples are - customer database, personnel database, inventory database and other databases containing data generated by business operations.
- b. **Management database:** These databases store data and information extracted from selected operational databases and external sources. They consist of summarized data and information which is mostly needed by organisational managers and other end users. Management databases are also called information databases.
- c. **Information warehouse databases:** An information warehouse stores data from current and previous years and such information is extracted ^(=collected) from various operational and management databases of an organisation. An important use of information warehouse databases is to determine the trend.
- d. **Distributed Databases:** These databases are generated by local work groups, departments at regional offices, branch offices, manufacturing plants and other work sites.
- e. **End User Databases:** These databases consist of a variety of data files developed by end users at their workstations. For example, users may store files generated by word processing packages or spreadsheet or other application packages or E-mails.
- f. **External Databases:** Organisations can access data from external, privately owned online databases or data banks, for a fee. Data is available in the form of statistics on economic trends.
- g. **Text Databases:** These databases will develop when computers are used to create and store documents electronically. Big corporations and government agencies have developed large text databases containing documents of all kinds.
- h. **Image Databases:** The above stated databases consist of data in traditional alphanumeric form. But in image databases a wide variety of images can also be stored electronically. Thousands of pages of business documents, such as customer correspondence, purchase orders, invoices, sales catalogues and service manuals can be scanned and stored as images on a single optical disk.

Q.No.24. Write about distributed databases? (N04 – 5M, N07 – 1M, RTP)

In a centralised system, the computer system and the database are both found at one location. Databases in this system are called centralised databases. If an organisation wants to decentralise its processing, the database or parts of the database must be distributed. There are basically two ways to distribute a database.

Replicated database: In this method duplicates of entire database are provided at all sites. This method is recommended if it is necessary for every location to have frequent access to the same data. However, this

method is very expensive and it is difficult to maintain consistency. Of course, replication provides some security to the database. If data at one particular location is lost, duplicate will be readily available.

Partitioned database: Another means of distributing the database is to partition it. In this method the database is divided into segments that are appropriate for certain locations and those segments are distributed only to those locations. Many organisations with many locations, partition the database hierarchically. For example, detailed data such as payroll and sales data are kept close to their source i.e. local site. Regional and national locations receive summaries of detailed data.

Advantages:

- » Distributed database reduce costs for an organisation because they reduce transfer of data between remote sites.
- » They can also reduce the response time for filing orders, answering customer requests or providing managers with information.
- » Immediate updation of data and faster communication of information within the firm.

Disadvantages:

- » Very difficult to control a distributed database system because it increases problems of security and need more computer resources.
- » If the distribution of database is not done carefully, many of the advantages of database will be lost.
- » Need more computer resources.
- » Increases data redundancy.

Q.No.25. Write about object oriented database. (M05 – 1M, RTP)

File and database structures were unable to meet the changing needs of application programs. When organisations recognized the need for integrated databases, hierarchical, network and subsequently relational data structures emerged.

Relational database structures are very flexible. But they are best suited to handle character and numeric data. The results of many engineering applications such as Computer Aided Design (CAD), Computer Aided Engineering (CAE) are not handled easily by relational systems. Expert Systems, Image Processing Systems, Multimedia Systems, Computer Aided Software Engineering (CASE), Project Management Software, Presentation Graphics Packages produce lot of important information which is not supported by traditional file and database structures. New structures are needed to handle such information.

Object Oriented Databases are new structures which can handle such information. Object Oriented Data Base (OODB) is a natural extension of OOP. Like other database management systems OODB systems can be either active or passive.

In an active OODB, every object is described by a set of attributes describing what the object is. The behaviour of the object is also included. For example, the object for a CAD design of a new product could include instructions for on-screen display, shrinking, exploding and rotation.

Object-oriented databases are still evolving, but the goal is to provide a repository of reusable objects that need to be created only once. This can greatly accelerate application development time and reduce application backlogs.

Q.No.26. Write about Client-server Database.

Meaning:

- » It is designed in such a way that one system can connect to another system to ask a question or instruct to perform a job.
- » The system that asks the questions and issues the instructions is the client and the system answering the queries and responding to the instructions is the server.
- » The client machine contains the user interface logic, business logic and the database logic and the server machine contains the database.
- » Both are interconnected through high bandwidth network.
- » The user interface program or front end program is called client and back end program is called server. Client and server computers can interact in an environment which is based on heterogeneous hardware / software platforms.

- » The server does all higher level functions which it alone can do leaving the client to perform lower level functions.
- » The system is scaleable i.e. any number of clients may be added or removed and the shared resources may be relocated to a larger and faster server or to multiple servers.

2-tier architecture:

- » The above is a 2-tier model which implies a complicated software distribution procedure.
- » All the application logic is executed on the personal computers. If a new software is released all such personal computers should be updated which is very costly, time consuming, complicated and error prone.
- » At the client end the software first has to be installed and then tested for correct execution.
- » Due to this it can't be assured that all clients work on the correct copy of the program.

3-tier & n-tier architecture:

- » 3-tier and n-tier client-server database designs try to solve these problems by simply transferring the application logic from the client back to the server.
- » It is achieved by inserting an application server tier between the data server tier and the client tier.
- » Client tier is responsible for data presentation, receiving user events and controlling the user interface.
- » The actual business logic is handled by the application server tier.
- » Business objects which implement the business rules, reside in application server tier which forms the central key to solve the 2-tier problems and protects the data from direct access by the clients.
- » Data server tier is responsible for data storage.

Q.No.27. Write about Knowledge Database.

- » A database system provides facilities to define, create, modify, delete and read data in a system.
- » Historically, the type of data maintained in a database system has been *declarative*, describing the static aspects of the real world objects and their associations. For e.g. a pay roll file and a personnel file can share data about salary of each and every employee, their positions, names, etc.
- » A database system can also be used to maintain *procedural data* describing the dynamic aspects of the real world objects and their associations. For e.g. a database can contain, several amended versions of enactments in the field of labour laws to facilitate management decisions in pay negotiations.
- » When both the *declarative* and *procedural* data are stored in a database it is called a knowledge database with more powerful data maintenance.

Q.No.28. Write about SQL? Explain its utility in DBMS. (N04 – 5M) (M06, M08 – 1M, RTP)

Query languages allow programmers, managers and other users to ask adhoc questions of the database without the aid ^(=help) of programmers. One form of query language is SQL or Structured Query Language. SQL is a set of about 30 English line commands that has become a standard in the database industry. Many vendors of software now provide SQL along with their database software.

SQL is used in many database management systems. Managers who understand SQL syntax can use the same set of commands in every database management system. Thus by learning SQL, manager can access data located in many database management systems.

The basic form of an SQL command is SELECT..... FROM... WHERE.....After SELECT the user lists the fields desired ^(=wanted) by him. After FROM, the name of the file or group of the records that contains those fields is listed. After WHERE any conditions for the search of the records are listed. For example, user may wish to SELECT all customer names FROM customer records WHERE the state in which the customer lives is Maharashtra. So, the following command will be entered.

```
SELECT      NAME, ADDRESS. CITY STATE, PIN
FROM        CUSTOMER
WHERE       STATE= 'Maharashtra'
```

This gives a list of names and addresses of all customers located in Maharashtra. Some query languages use natural language set of commands. These query languages are structured so that the command used is close to Standard English as possible. For example:

PRINT THE NAMES AND ADDRESSES OF ALL CUSTOMERS WHO LIVE IN MAHARASHTRA.

On the other hand managers need not learn COBOL, BASIC or other standard programming languages to access the database. It is sufficient to learn SQL.

Q.No.29. Write about different concepts of data storage?

Storage structure: For designing a storage structure and for storing data on some storage device the following decisions are to be taken:

- » How to make it linear,
- » How to partition the data structure so that it can be stored on some device.

File library: The file library function acts as an aspect of Operations Management Controls. It takes the responsibility of machine readable storage media. In this context, following four functions are undertaken:

- » Ensuring that removable storage media are stored in a secure and clean environment,
- » Ensuring that storage media are used for authorized purposes only,
- » Ensuring maintenance of storage media in good working condition and
- » Ensuring location of storage media at on-site / off-site facilities.

Importance of storage media:

- » A storage media can contain critical files and hence should be housed securely.
- » In case of mainframe operations, a separate room may be required to house storage media required during daily operations, restricting access to the room and fixing the responsibility in the hands of a file librarian.
- » Similar arrangements may be followed in case of off-site backup storage.
- » Both on-site and off-site facilities should be maintained in a constant temperature and in a dust free environment.

Managing bulk of removable storage media: For managing a bulk of removable storage media in an effective way an automated library system is needed which records the following –

- » An identifier of each storage medium,
- » Location where each storage medium is currently placed,
- » Identity of the person responsible for storage medium,
- » Identity of the present custodian of the storage medium,
- » List of files stored on each medium,
- » List of authorized persons who can access,
- » Date of purchase, history of use, history of difficulties in using the storage medium,
- » Date of expiry i.e. when the contents of the storage medium can be deleted,
- » Date of last release from and date of last return to the file library,

Care required to be taken in using storage media:

- » Use of storage media should be controlled with great care.
- » In mainframe environment, file librarians issue removable storage media as per authorized production schedule.
- » In other cases, file librarians issue it on the basis of an authorized requisition, only after recording the use of removable storage media.
- » In micro computer environments, the extent of control depends upon the criticality of data.
- » At off-site location, asset safeguarding and data integrity objectives may be seriously compromised.
- » Generally off-site personnel work for a vendor who may specialize in providing backup facilities. Off-site personnel may not have the knowledge of the persons who are authorized to access storage media.
- » Care should be taken when multiple files are stored in a single storage medium. When a storage medium is granted for accessing a particular file then it may also contain some other sensitive files.
- » It is better to store sensitive files in a separate storage medium.
- » Operating System should be designed to exercise control over an application program to restrict access to authorised files only.
- » Some sensitive data may get exposed to unauthorised persons. To avoid this, sensitive files should be removed out of storage medium as soon as their retention date expires.
- » Such media can then be reused to store other files, thereby reducing inventory requirement.

Care required to be taken to maintain the reliability of storage media:

- » Reliability of storage media should be monitored, especially in case of critical files, because reliability decreases with age. For e.g., several read / write errors occur in case of magnetic tapes and diskettes as they become older and used for several times.
- » Storage media should not be kept idle for long period of time, otherwise the possibility of read / write errors increases.
- » Even if it is necessary to backup the data for a longer period of time, it should be retrieved at regular intervals and backup files should be re-written to another fresh medium.

How to discard storage media?

- » When storage media becomes unreliable it is better to discard it, after removing all sensitive data.
- » Simply deleting the files on the storage media will not serve the purpose.
- » It may be necessary to carry out bulk erasure followed by writing random bits to the media followed by bulk erasure once again.
- » In some cases, unreliable storage media can be repaired. For e.g. magnetic tapes can be cleaned to improve its reliability and to enhance its life.
- » If storage media is required to be sent outside the organization for cleaning or repair, care should be taken to erase any sensitive data. For e.g. data on magnetic tapes can be erased by degaussing or demagnetization.

Where to store removable storage media?

- » Removable storage media should be located on-site when they are required to support some online application systems.
- » They should be located off-site when intended for backup and recovery.
- » In a mainframe environment, file librarian is responsible to manage transport of removable storage media to and from off-site locations. These movements should comply with the backup schedules.
- » Backup schedules are prepared by a team consisting of security administrator, data administrator, application project manager, system development manager, operations manager and file librarian.
- » Backup is prepared for minicomputer disk storage, local area network file server and critical microcomputer files.

Q.No.30. State the documents that are required to support a system in an organisation?

The documentations that are needed to support a system in an organization are:

Main documents	Ancillary documents
a. Strategic and operational plans, b. Application systems and program documentation, c. System software and utility program documentation, d. Database documentation, e. Operations manuals, f. User manuals, g. Standard manuals.	B memoranda books and journals

How to store this documentation?

- » Generally these are kept in an automated form.
- » Some software provides documentation on optical discs.
- » However, most of the documentation is still kept in hard copy format since it is more advantageous than online documentation.

Q.No.31. Write down the steps to be taken for proper system documentation? (Or) System documentation is not escaped from difficulty. Explain.**Difficulties in system's documentation:**

- a. Responsibility for documentation is spread throughout the organization. For example, a librarian may be responsible for documentation supporting mainframe and mini computer systems. On the other hand, documentation supporting micro computer system may be the responsibility of its users,

- b. Documentation is maintained in multiple forms and in multiple locations. For example, some part may exist in magnetic form, some other part in hard copy form and some other part in microfilm form,
- c. Because of the density and dispersion of documentation it is very difficult to maintain, update, access and take backup.

Responsibilities of documentation librarian: The responsibilities of documentation librarians are to ensure that

- a. Documentation is stored securely,
- b. Only authorized users can have access to it,
- c. It is updated,
- d. Adequate backup exists for it.

Problems of improper documentation: Many organisations acquire a large number of software packages to support their microcomputer operations. If the software inventory is not properly managed by the documentation librarian then it may lead to any of the following problems:

- a. Purchase of too many copies of the software,
- b. Loss or theft of the software,
- c. Loss or theft of the documentation,
- d. Illegal copying of the software,
- e. Use of software not complying with the terms and conditions of the software license,
- f. Absence of the software backup or improper backup.

Solution to these problems: Several types of software are available to minimize the above stated difficulties by taking the responsibility of maintaining the records of purchases, distributions and uses of the software and its related documentation and ensuring compliance with terms and conditions of the licensing agreements by the users. For example, some local area network Operation Systems can provide a utility which can generate a list of all software located in workstations or file servers in the network.

Q.No.32. Write short notes on Program Library Management System Software. (N02, N04 – 5M)

Meaning: Program library management system is a database of system and application software used by the organisation. It may include

- a) Application and system software program code,
- b) **Job control statements** that identify the following which manage the processing:
 - » resources used,
 - » processes to be performed, and
 - » processing parameters.

Capabilities / Functions of PLMS:

- a. It provides several functional capabilities to effectively and efficiently manage **data center** software inventory which include:
 - » Application Program Code,
 - » System Software Code,
 - » Job Control Statements.
- b. It possesses integrity capabilities such that for each source program
 - » A modification number is assigned,
 - » A version number is assigned,
 - » And it is associated with a creation date.
- c. It uses:
 - » Password,
 - » Encryption,
 - » Data Compression,
 - » Automatic backup
- d. It possesses update capabilities with the facilities of:
 - » Addition,
 - » Modification,
 - » Deletion,
 - » Re-sequencing library members
- e. It possesses reporting capabilities for being reviewed by management and end users by preparing lists of:
 - » Additions,
 - » Deletions,
 - » Modifications,
 - » Library catalogue,
 - » Library members attributes
- f. It possesses interface capabilities with the:
 - » Operating System,
 - » Job scheduling system,
 - » Access control system,
 - » Online program management
- g. It controls movement of programs from test to production status.
- h. At last, it changes controls to application programs.

Q.No.33. Define user interface. What are the elements to be considered in designing user interface?

Design of user interface: User interface involves designing of the ways in which users will interact with the system. Following elements should be considered in designing user interface:

- » Source documents to capture data,
- » Hard copy output reports,
- » Screen layout for input of source documents,
- » Enquiry screens for database queries,
- » Command languages for Decision Support Systems,
- » Query languages for the database,
- » Graphic display and color or non-monochromatic display,
- » Voice output to answer users or answer queries,
- » Screen layout manipulation by mouse or light pen,
- » Icons for representation of the output.

Development of User interface: User interface design is developed as follows

- » Identifying system users and classifying them into homogeneous groups,
- » Understanding the characteristics of the user group. For example, whether the system is handled by novice users or experts,
- » Bring out the tasks which the users wish to perform using the system,
- » Commencing a preliminary design of the form of interaction that will support these tasks. Generally, prototyping tools are employed to refine the design aspect with the users.

Q.No.34. Explain the terms (a) Backup (b) Recovery (c) Disaster recovery.**Backup:**

- » 'Backup' is a utility program used to make a copy of the contents of database files and log files.
- » These additional copies may be used to restore the original files after the event of a data loss. These additional copies are called 'backup files'.
- » The database files consists of a database root file, log file, mirror log file and other database files called dbspaces.
- » Generally 'backup and recovery' is treated as one topic and 'disaster recovery' as another.

Recovery:

- » '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, disk drive, controller card or power supply.
- » Media failure is the result of unexpected database error while processing data.
- » Before beginning database recovery, it is always advisable to have a back up copy of the failed database. This preserves the existing situation.

Disaster recovery:

- » Disaster recovery is the process of regaining access to the data, hardware and software necessary to resume critical business operations after a natural or human induced disaster.
- » 'Disaster recovery' is different from database recovery because the operating system and all the related software must be recovered before any database recovery can begin.

Q.No.35. Write about database files that make up a database?

Databases consist of disk files that store data. When we create a database, a main database file or root file is created. This main database file contains database tables, system tables and indexes. The database files consist of a database root file, log file, mirror log file and other database files called dbspaces.

A **transaction log** is a file that records database modifications. Database modifications consist of insertions, updations, deletions, commits, rollbacks and database schema changes. A transaction log is not compulsory, but it is recommended. (M08 – 1M)

A **mirror log** is an optional file and has a file extension of **.mlg**. It is a copy of the transaction log and provides additional security against the loss of data and it is very much useful when the transaction log becomes unusable.

Q.No.36. Write about different types of backup?

Database backups can be performed while the database is being actively accessed (online) or when the database is shutdown (offline).

Online backup:

- » Database backup which is performed while the database is being actively accessed is called on-line backup.
- » An online database backup can be performed by executing the command-line or from the 'Backup Database' utility.
- » 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 the transactions recorded since the last database backup.

Off line backup:

- » Performance of backup when the database is shutdown is called offline backup.
- » When a database goes through a normal shutdown process (i.e. the process is not being cancelled) the database engine makes permanent changes in the database files.

Live backup:

- » A live backup provides a redundant copy of the transaction log for restarting system on a secondary machine when the primary database server machine becomes unusable.
- » A live backup is carried out by using the BACKUP utility with the command line option.

Full and incremental database backup:

- » A database backup can be either full or incremental.
- » For a full backup, 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 an incremental backup is performed the mirror log is not backed up.
- » When we backup and rename the log files the transaction and mirror log file are renamed and new log files are created.
- » We must plan to manually back up the mirror log. Be aware of this when planning out backup and recovery strategy.

Q.No.37. Write about different steps in the development of a back up and recovery strategy?

Following are the steps in the development of a back up and recovery strategy:

1. **Understand what backup and recovery means to the business:** How long can a business survive without access to the corporate data? Express the answer in terms of minutes, hours or days.

If the recovery time is in minutes then database backup and recovery is critical to the business and it is very important to implement some kind of backup and recovery strategy. If recovery time can take hours then we can have more time to perform the tasks. If recovery can be expressed in terms of days then time appears to be less important.
2. **Management commits time and resources for the project:** Management must commit some financial resources towards the development and implementation of a backup and recovery strategy. The strategy can be basic or quite extensive depending upon the business needs of the company. After developing a backup and recovery strategy, management should be informed of the expected backup and recovery times. Anticipate management countering the timings by preparing alternative solutions. Then, management will decide the best solution.
3. **Develop, test, time, document, health check, deploy and monitor:** These are the most important phases in the development of a backup and recovery strategy:
 - ❖ Create backup and recovery commands. Verify whether these commands produce the desired results.

- ❖ Prepare time estimates for executing backup and recovery commands. Use this information to identify what commands will be executed and when.
 - ❖ Document the backup commands and create written procedures for outlining
 - where the backups are kept,
 - identify the naming convention used and the kind of backup performed.
 - ❖ Incorporate health checks into the backup procedures to ensure that the database is not corrupt.
 - ❖ Deployment of backup and recovery consists of setting up of backup procedures on the production server. Provide necessary hardware and any other supporting software to perform these tasks.
 - ❖ Monitor backup procedures to avoid unexpected errors. If any changes take place in the process, reflect those changes in the documentation.
4. **Beware of external factors that affect recovery:** External factors that affect database recovery are time, hardware and software. Allow additional recovery time for performing miscellaneous tasks such as entering recovery commands or retrieving and loading tapes. Factors that influence time are size of database files, recovery medium, disk space and unexpected errors. As the backup and recovery strategy develops it is necessary to check the performance of the equipment and software to ensure that it meets our expectations.
5. **Protect database backups by performing health checks:** Database health checks are run against the database and log files to ensure that they are not corrupt. If the database file is found to be corrupt, we need to recover it from the previous database backup.

Q.No.38. Define the term data warehouse. (M08 – 5M) What is the need for the development of data warehouses?

A data warehouse is a computer database that collects, integrates and stores an organization's data with the aim of producing accurate and timely management information and supporting data analysis. Data warehouse has become a distinct type of computer database during the late 1980's and early 1990's.

Need for data warehouses: They are developed to meet the growing demand for management information and analysis that could not be met by operational systems. Operational systems were unable to meet this need because of following reasons:

- » The processing load of reporting has reduced the response time of the operational systems,
- » The database designs of operational systems were not optimised for information analysis and reporting,
- » Most organisations had more than one operational system, so company wide reporting could not be supported from a single system, and
- » To develop reports in operational systems it was required to write specific computer programs which was very slow and expensive.

As a result, separate computer databases are built that are specifically designed to support management information and for the purpose of analysis.

These data warehouses were able to bring in data from different data sources such as mainframe computers, minicomputers, personal computers and office automation software such as spreadsheets and integrate this information in a single place. This capability is coupled with user-friendly reporting tools.

Q.No.39. What are the various stages in the development of data warehouses?

As technology improved and user requirements increased, data warehouses have evolved through several fundamental stages:

- a. **Offline Operational Databases:** Data warehouses in this initial stage are developed by simply copying the database of an operational system to an off-line server. As a result the processing load of reporting does not impact the performance of the operational system.
- b. **Offline data warehouse:** Data warehouses in this stage are updated on a regular time cycle (usually daily, weekly or monthly) from the operational systems and the data is stored in an integrated reporting oriented data structure.
- c. **Real time data warehouse:** Data warehouses at this stage are updated on a transaction or event basis i.e. data warehouse is updated every time an operational system performs a transaction.

- d. **Integrated data warehouse:** Data warehouses at this stage are used to generate activity or transactions that are passed back into the operational systems for use in the daily activity of the organization.

Q.No.40. Write about various components of a data warehouse? (RTP)

The primary components of the majority of data warehouses are:

1. **Data Sources:** Data sources refer to any electronic repository of information that contains data of interest for management use or analysis. This definition covers mainframe databases, client-server databases, PC databases, spreadsheets and any other electronic store of data. Data needs to be passed from these systems to the data warehouse either on a transaction-by-transaction basis (real-time data warehouses) or on a regular cycle, e.g. daily or weekly. (off line data warehouses).
2. **Data Transformation:** The Data Transformation layer receives data from the data sources, cleans it and standardises it and loads it into data repository. Generally this is called "staging of data" because data generally passes through a temporary database while it is being transformed. Transformation of data can be done either manually or through a specific type of software. Regardless of the software used, following activities occur during data transformation:
 - » Comparing data from different systems to improve data quality (For e.g. Date of birth for a customer may be blank in one system but contain valid data in another system. The data warehouse will retain the date of birth field from the second system)
 - » Standardising data and codes.
 - » Integrating data from different systems (e.g. if one system stores orders and another stores customers, then these data elements need to be linked together)
 - » Performing other functions such as determining change to be made to files to reduce data load times, generating or finding replacement keys for data, etc.
3. **Data Warehouse:** The data warehouse is a relational database organised to hold information in a structure that best supports reporting and analysis. Most data warehouses hold information for at least 1 year. As a result these databases can become very large.
4. **Reporting:** If the data warehouse is to be useful, the data in the data warehouse must be available to the organisation's staff. Several software applications can perform this function or reporting can be custom developed. Examples of different types of reporting tools are:
 - a. **Business intelligence tools:** These are software applications that simplify the process of development and production of business reports based on data warehouse data.
 - b. **Executive information systems:** These are software applications that are used to display complex business metrics and information in a graphical way to allow rapid understanding.
 - c. **OLAP tools:** OLAP tools arrange data into logical multi dimensional structures and users can select the dimension in which he wants to view data.
 - d. **Data Mining:** Data mining tools allow users to perform detailed mathematical and statistical calculations to detect trends, identify patterns and analyse data.
5. **Metadata:** Metadata or "data about data" is used to inform operators and users of the data warehouse about its status and the nature of information held in the data warehouse. For example, the most recent data load date, the business meaning of a data item, the number of users that are currently logged in etc.
6. **Operations:** Data warehouse operations consist of loading, manipulating and extracting data from data warehouse. Operations also cover functions like user management, security, capacity management, etc.

Optional components: In addition, the following components also exist in some data warehouses:

- a. **Dependent data marts:** It is a physical database that receives all its information from the data warehouse. The main purpose is to provide a sub-set of the data warehouse's data for a specific purpose or to a specific sub-group of the organisation.
- b. **Logical data marts:** It is a filtered view of the main data warehouse but does not physically exist as a separate data copy. The main advantage of this approach is that it does not require additional disk space.
- c. **Operational data store:** It is an integrated database of operational data. It contains current or near term data. An ODS may contain 30 to 60 days of information, while a data warehouse typically contains years of data.

ODS's are used to provide some near real time reporting which a data warehouse can't provide. This is so because Data Warehouse's loading time and architecture prevents it from providing near real time reporting.

Q.No.41. Write about different methods of storing data in a data warehouse?

All data warehouses store their data grouped together by subject areas that reflect the general usage of the data (Customer, Product, Finance etc.) The general principle used in the majority of data warehouses is that data is stored at its most elemental level for use in reporting and information analysis. Within this generic intent, there are two primary approaches to organise data in a data warehouse.

Dimensional approach:

a. Meaning:

- » In this approach, information is stored as "facts" which are numeric or text data that capture specific data about a single transaction or event and "dimensions" which contain reference information that allow each transaction or event to be classified in various ways.
- » As an example, a sales transaction would be broken up into facts such as the number of products ordered, the price paid and dimensions such as date, customer, product, geographical location and sales person.

b. Advantages:

- » The Data Warehouse is easy for business staff with limited information technology experience to understand and use.
- » As the data is pre-processed into the dimensional form, the Data Warehouse tends to operate very quickly.

c. Disadvantage: It is quite difficult to add or change later if the company changes the way in which it does business.

Normalisation approach:

- » In this approach, the data in the data warehouse is stored in third normal form.
- » The main advantage of this approach is that it is quite straightforward to add new information into the database.
- » The primary disadvantage of this approach is that it is very slow to produce information and reports.

Q.No.42. List out the advantages and concerns in using data warehouses.

Advantages of using data warehouses: Following are some of the advantages in using a data warehouse:

- » Enhances end-user access to a wide variety of data.
- » Increases data consistency.
- » Increases productivity and decreases computing costs.
- » It is possible to combine data from different sources, in one place.
- » It provides an infrastructure that could support changes to data and replication of the changed data back into the operational systems.

Concerns in using data warehouses:

- » Extracting, cleaning and loading data could be time consuming.
- » Data warehousing project scope might increase.
- » Compatibility problems with systems already in use e.g. transaction processing system.
- » Providing training to end-users, who are unable to use data warehouse.
- » Security could be a major issue, especially if the data warehouse is web accessible.
- » A data warehouse is a high maintenance system.

Q.No.43. Write about data mining and various steps in data mining?

Data Mining:

- » Data mining is concerned with the analysis of data and the use of software techniques for finding patterns and regularities in sets of data.
- » Here a computer software is responsible for finding the patterns by identifying the underlying rules and features in the data.

- » The logic behind data mining is that it is possible to strike gold in unexpected places.
- » Data mining software extracts patterns that are previously unrecognized or so obvious that no one has noticed them before.
- » The analysis process starts with a set of data, uses a methodology to develop an optimal representation of the data during which time knowledge is acquired.
- » Once knowledge is acquired this can be extended to larger sets of data basing on the assumption that the larger data set has a structure which is similar to the sample data.
- » This process is similar to mining activity where large amounts of low grade materials are sifted through in order to find some valuable thing.

Steps in Data Mining: The data mining activity starts with raw data and ends with extracted knowledge which is acquired from the following stages:

- a. **Selection:** Selecting or segmenting the data according to some criteria, e.g. all those people who own a car, is the first step in data mining. In this way subsets of data can be determined.
- b. **Preprocessing:** This is the data cleansing stage where certain information is removed which is deemed unnecessary and may slow down the queries. For example, it is unnecessary to study the sex of a patient when studying pregnancy. As data is drawn from several sources, there is a possibility of inconsistency. So, data is reconfigured to ensure consistency. For example sex of a person may be recorded as 'F' or 'M' in some cases and in some other cases it may be recorded as 1 or 0. So, data should be reconfigured to ensure consistency.
- c. **Transformation:** The data is not merely transferred across but transformed in order to be suitable for the task of data mining. The data is made useable and navigable.
- d. **Data mining:** This stage is concerned with the extraction of patterns from the data. A pattern can be defined as a given set of facts (data) F , a language L , and some measure of certainty C . A pattern is a statement S in L that describes relationships among a subset F_s of F with a certainty C such that S is simpler in some sense than the enumeration of all the facts in F_s .
- e. **Interpretation and evaluation:** The patterns identified by the system are interpreted into knowledge which can then be used to support human decision-making e.g. prediction and classification tasks, summarizing the contents of a database or explaining observed phenomena.

A category: Q.NO: 7, 8, 13, 14, 15, 17, 20, 22, 23, 24, 25, 27, 28, 30, 31, 32, 34, 36, 37, 38, 40, 42, 43

B category: Q.NO: 6, 21, 26, 35, 39

C category: Q.NO: 1, 2, 3, 4, 5, 9, 10, 11, 12, 16, 19, 29, 33, 41

1 Mark Questions:

ASCII	Primary key	Secondary key	Referential integrity
Data	Information		
	Record	Data File	Database
Sequential Access	Direct Access	DDL	DML
Data Independence	Data Dictionary / Directory	DBMS	Operational Database
Management Database	Information warehouse Database	Distributed Database	End user Database
External Database	Text Database	Image Database	Object Oriented Database
Client-Server Database	Knowledge Database	SQL	Backup
Recovery	Disaster Recovery	Transaction Log	Mirror Log
Online Backup	Offline Backup	Data warehouse	Data Transformation
Metadata	Data Mining		

The End