BASIC CONCEPT

- **DATA**

Data is called as distinct pieces of information. Data is isolated facts.
e.g. Student number, Student name, Address etc. is defined as data or information about student.

- **INFORMATION**

Any meaningful data is called as information.

- **FIELD**

Any particular column or data item from a record is defined as a field. In other words a vertical line from a database is called as a field.
Each individual item of data stored in a record is known as a field. We have arranged the information under different columns; we also store the information in record under different columns. These columns in DBMS terminology are called fields.

e.g. Suppose there is a table named Student having information like
StudNo
StudName
Address
Here StudNo is called as one field, StudName is the other field and so on.

- **RECORD**

Complete information about any particular person or subject is called as a record. In other words, a horizontal line from a database is defined as a record.
A record is a collection of similar type of data item treated as a single unit. e.g. in address database data about a particular person, such as Name, Address, PIN code are the individual record. So the physical (horizontal) line in the database file is record.

e.g. Suppose there is a table named Student having fields like
StudNo
StudName
Address
Here StudNo, StudName, Address about any particular student is called as a record.

- **DATABASE**

A database is a collection of related information about particular subject or person.
e.g. Suppose there is a table named Student having fields like

StudNo
StudName
Address

Here Student number, student name and address is a related information about student and therefore a collection of these information about all students is called as database.
The database is an *organized* collection *related* data. Here *organized* word is stressed because only organized information is a database. Any unorganized information is just dump or a pile and it cannot be called a database.
Database can be defined as a collection of inter-related relevant data stored together to serve multiple applications.

- **DATABASE FILE**

Information stored in a computer file using database management system is called as database file.
A file is *collection* of *similar* groups of data.

- **DBMS (DATABASE MANAGEMENT SYSTEM)**

A DBMS is a collection of software packages for processing the databases. In other words, a database on which a different processes are done then that system is defined as a database management system.
DBMS is the software that function as the interface between users, other application program and the database itself. It allows the data to be stored, maintained, manipulated and retrieved.

The processes can be
- Adding records
- Modifying records
- Deleting records
- Arranging records
- Searching records
- Printing records

- **PRIMARY KEY**

The field that uniquely identifies each record from a table (database file) is called as primary key. Primary key must not be null and must be unique across the column. A primary key is used to relate a table to foreign keys in other tables.
e.g. Suppose there is a table named Student having fields like

StudNo, StudName, Address

Here StudNo must be unique for all students so we must define StudNo as primary key field.

- **FOREIGN KEY**

Foreign key represents relationships between tables. A foreign key is a column or group of columns which values are derived from the primary key table.

The existing of a foreign key implies that the table with the foreign key is related to the primary key table from which the foreign key is derived. A foreign key must have a corresponding primary key value in the primary key table to have a meaning.

e.g. Suppose there is a table named Student having fields like

StudNo, StudName, Address
And there is another table named Result having fields like
StudNo, ITM_201, ITM_202,

Here StudNo from Result table is defined as foreign key. So it must match with the field name and values of StudNo from Student table. Here we can define relationship and combine information between Student table and Result table using field StudNo.

- **DRAWBACKS OF TRADITIONAL FILE**

The following are the drawbacks of traditional file.

i. Data Redundancy

- Many applications used their own special files of data.
- So some data items may common to several applications.
- For example, in a bank same customer name might appear in checking account file, a saving account file and an installment loan file.
- The customer is same but the related field often had a different name in the various account file.
- The field name may be CNAME in checking account file, SNAME in saving account file and INAME in installment loan file.
- The same field might have different length in the various files.
- CNAME could be up to 20 characters, SNAME and INAME might be limited to 15 characters.
- So redundancy increased the overhead cost of maintenance and storage.
- Data redundancy also increased the risk of inconsistency among the various versions of common data.

ii. Poor Data Control

- In the traditional file there is no centralizing control of data.
- There is always a chance that the various department of a company will be inconsistent in their terminology.
- I.e. It may possible that same field name has different meaning in a company.
- For example, in a bank one meaning of account is when applied to savings and meaning is different when applied to loans.
- It may also possible that different field name had same meaning in a company.
- For example, the meaning of policy and case is same for insurance company.

iii. Insufficient Data Manipulation Capabilities

- Suppose there is a table named Student having fields like StudNo, StudName and Address.
- If we want to access a particular record we can access that record using the key StudNo.
- And suppose there is another table named Result having fields like StudNo, Subject1 and Subject2.
- Now if we want a set of related record from both tables, such information would be difficult for traditional file.
- Because here file systems are unable to provide strong connections between data in different files.

iv. Excessive Programming Effort

- A new application program often required an entirely new set of file definitions.
- Even though an existing file may contain some of the data needed, the application often required a number of other data items.

- As a result, the programmer had to recode definitions of needed data items from the existing file as well as definitions of all new data items
- Thus there was a heavy interdependence between program and data.

- **ADVANTAGES OF DBMS**

The following are the advantages of DBMS.

i. Reduction of Redundancy

- In DBMS there is a centralize control of data.
- So it avoids unnecessary duplication of data and reduces the total amount of data storage required.
- It also eliminates the extra processing necessary to trace the required data in a large mass of data.
- As DBMS avoids duplication, it eliminates the inconsistencies that present in redundant data.
- Any redundancies that exist in the DBMS are controlled.

ii. Shared Data

- A database allows the sharing of data under its control by any number of application program or users.

iii. Integrity

- DBMS provides data integrity.
- Data integrity means that the data contained in the database is both accurate and consistent.
- Therefore, data values being entered for storage could be checked to ensure that they fall within a specified range and are of the correct format.
- For example, the value for the age of and employee may be in the range of 16 and 75.

iv. Security

- Data may be importance to an organization and may be confidential. Such confidential data must not be accessed by unauthorized persons.
- The DBA who has the responsibility for the data in the DBMS can ensure proper schemes for access to the DBMS and additional checks before permitting access to sensitive data.
- Different levels of security could be implemented for various types of data and operations.

v. Conflict Resolution

- Since the database is under the control of DBA, she or he should resolve the conflicting requirements of various users and applications.
- The DBA chooses the best file structure and access method to get optimal performance for the response critical applications, while permitting the critical applications to continue to use the database.

vi. Data Independence

- Data independence considered from two points of view :

Physical Data Independence

- It allows changes in physical storage devices or organization of the files to be made without requiring change in the conceptual view or any of the external view in the programs using the database. The file structure may change without any need for changes in the application programs.

Logical Data Independence

- It implies that application programs need not be changed if fields are added to and existing record nor do they have to be changed if fields not used by application programs are deleted.
- Thus data independence allows for changes at one level of the database without affecting other level.

- **DISADVANTAGES OF DBMS**

The following are the disadvantages of DBMS.

i. Cost

- The cost of purchasing or developing the software.
- Hardware has to be upgraded to allow for the program and the work spaces required for their execution and storage

ii. Complexity of Backup/Recovery

- As database is centralized, it requires that in case of failure the data we should take backup.
- So here backup/recovery operations are complex.

iii. Problems with centralization

- Centralization also means that the data is accessible from a single source so in case of downtimes or failures the operation of the organization will be disturb.

- **ORGANIZATION OF DBMS**

OR

THREE–LEVEL ARCHITECTURE

The architecture is divided into three levels : External level,
Conceptual level and
Internal level.

- The view at each of these levels is described by a scheme.
- A scheme is an outline or a plan that describes the records and relationships existing in the view.
- The word scheme, which means a systematic plan for attaining some goal, is used in the database with the word schema.

External or User View

- External view is described by means of a scheme called an external schema.
- The external schema consists of the definition of the logical records and the relationships in the external view.
- It also contains the method of deriving the objects in the external view from the objects in the conceptual view.
- The objects includes entities, attributes and relationships.

Conceptual or Global View

- All this level all the database entities and the relationships among them are included.
- This conceptual view is defined by the conceptual schema.
- There is only one conceptual schema per database. This schema also contains the method of deriving the objects in the conceptual view from the objects in the internal view.
- It also include features that specify the checks to retain data consistency and integrity.

Internal View

- It indicates how the data will be stored and described the data structures and access methods to be used by the database.
- The internal view is expressed by the internal schema.

• COMPONENTS OF DBMS

1. Classification of DBMS Users

The users of database system is classified in the following groups.

i. Naïve Users

- Users who need not be aware of the presence of the database system or any other system supporting their usage are considered naïve users.
- A user of an automation falls in this category.

ii. Online Users

- These are users who may communicate with the database directly via an online terminal or indirectly via a user interface and application program.
- These users are aware of the presence of the database system.

iii. Application Programmers

- Professional programmers who are responsible for developing application programs or user interfaces utilized by the naïve users fall into this category.
- The application programs could be written in the general purpose programming language such as Assembler, C, COBOL, FORTRAN, PASCAL and includes the commands to manipulate the database.

iv. Database Administrator

- Centralized control of the database is done by a person or group of persons under the database administrator (DBA).
- They are the users who are most familiar with the database and are responsible for creating, modifying and maintaining its three levels.

2. DBMS Facilities

Two main facilities are provided by a DBMS.

i. Data Definition Facility or Data Definition Language (DDL)

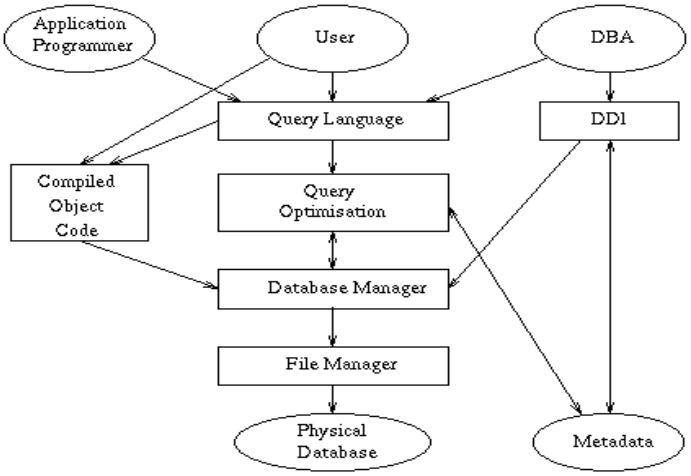
- DBMS provide a facility known as data definition language which can be used to define the conceptual scheme and also give some details about how to implement this scheme in the physical devices used to store the data.
- This definition includes all the entity sets and their associated attributes as well as the relationships among the entity sets.
- It also includes any constraints that have to be maintained including the constraints on the value that can be assigned to a given attribute and the constraints on the values assigned to different attributes in the same of different records.

ii. Data Manipulation Facility or Data Manipulation Language (DML)

- The language used to manipulate data in the database is called data manipulation language.

– Data manipulation involves retrieval of data from the database, insertion of new data into the database and deletion or modification of existing data.
–

- **STRUCTURE OF DBMS**



iii. Data Definition Language Compiler

– The DDL compiler converts the data definition statements into a set of tables.
– The tables contains the data concerning the database and are in a form that can be used by other components of the DBMS.

iv. Data Manager

– The data manager is the central software component of the DBMS.
– One of the function of the data manager is to convert operations in the user's queries coming directly via query processor or indirectly via an application program.
– It also interface with the file system
– It also maintains consistency and integrity of the data as well as also security.

v. File Manager

– Responsibility for the file structure of the files and managing the file space rests with the file manager.
– It also responsible for locating the block containing the required record, requesting this block from the disk manager and transmitting the required record to the data manager.

vi. Disk Manager

– The disk manager is part of the operating system of the host computer and all physical input and output operations are performed by it. The disk manager transfers the block or page requested by the file manager so that the latter need not be concerned with the physical characteristics of the storage media

vii. Query Processor

- The query processor is used to interpret the online user's query and convert it into an efficient series of operations in a form capable of being sent to the data manager for execution.

viii. Telecommunication System

- Online users of computer system, whether remote or local, communicate with it by sending and receiving messages over communication lines.
- These messages are routed via an independent software system called a telecommunication system.

ix. Data Files

- Data files contain the data portion of the database.

x. Data Dictionary

- Information of the structure and usage of data contained in the database is maintained by data dictionary.
- Each database user can consult the data dictionary to learn what each piece of data and the various synonyms of the data fields mean.
- In an integrated system (i.e. system where the data dictionary is part of the DBMS) the data dictionary stores the information regarding the source of each data field value, the frequency of its use, updates information including who and when of each update

- **DATABASE LIFE CYCLE**

Database life cycle consists of the following stages.

xi. Requirement Analysis

- Requirement involves the scope of the database identifying management and functional area information requirements and establishing hardware/software requirements.
- Information requirements are determined from questionnaire responses, interviews with managers, clerical users and reports and forms currently being used.

xii. Data Modeling

- Data model is a collection of conceptual tools used for describing data, data relationships, data semantics and data constraints.

xiii. Implementation

- Implementation may not be a creative process but certainly is a difficult task.
- This is because users have to accept the system. Hence human considerations will have to be attended very carefully.
- Users' training and availability of users' reference manuals with procedures to tackle trouble shooting are a must.

xiv. Testing

- Testing procedures are needed to ensure that the database system fulfills user requirements and operates without major problems.

xv. Maintenance

- This is the last step in the database life cycle.

- It takes the longest duration. Maintenance may be corrective, adaptive or perfective.
- In corrective maintenance errors or bugs are rectified.
- In the adaptive maintenance the user requirements if any are still considered and the necessary changes are made.
- In perfective maintenance efforts will be constantly going on to perfect the system in terms of response time and resource requirements.