DBMS

A database is a collection of related information stored so that it is available to many users for different purposes. The content of database is obtained by combining data from all the different sources in an organization. So that data are available to all users and redundant can be eliminated or at least minimize.

A computer database gives us some electronic filing system which has a large number of ways of cross-referencing and this allows the user many different ways in which to reorganize and retrieve data. The management of data in a database system is done by means of a general purpose software package called a database management system. A database management system, therefore, is a combination of hardware and software that can be used to set up and monitor a database, and can manage the updating and retrieval of database that has been stored in it. Some commercially available DBMS are INGRESS, ORACLE, Sybase. Most database management systems have the following facilities/capabilities.

- Creating a file, addition of data, deletion of data, modification of data, creation addition and deletion of entire file.
- Retrieving data collectively or selectively.
- The data stored can be sorted or indexed at the user's direction. Various reports can be produced from the system.

Advantage of DBMS:-

1. **<u>Reduction of redundancies</u>**- Centralized control of database by the DBA avoids unnecessary duplication of data and effectively reduces the total amount of data storage required. It also eliminated the extra necessary to trace the required data in a large mass of data. Another advantage of avoiding duplication is the elimination of the inconsistencies that tend to be present in redundant

data files. Any redundancies that exists in the DBMS are controlled and the system ensures that these multiple copies are consistent.

2. **Data Sharing**:- A database allows the sharing of data under its control by any number of application programs or users.
3. **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. Another integrity check that could be incorporated in the database is to ensure that if there is a reference to certain object, that object must exist.
4. **Data security:-** Data is of vital importance to an organization and may be confidential. Such confidential data must not be accessed by unauthorized persons. The DBA who has the ultimate responsibility for the data in the DBMS can ensure that proper access procedures are followed, including proper authentication schemas for access to the DBMS and additional checks before permitting access to sensitive data.
5. **Conflict resolution:-** Since the database is under the control of the DBA, she or he should resolve the conflicting requirements of various users and applications. In essence, the DBA chooses the best file structure and access method to get optimal performance for the response- critical application, while permitting less critical application to continue to use the database, albeit with a relatively slower response.
6. **Data Independence:-** Data independence, is usually considered from two points of view:

physical data independence and logical data independence.

Physical data independence allows changes in the physical storage device organization of the files to be made without requiring changes in the conceptual vie or any of the external views and hence in the application programs using the database. Thus the file may migrate from one type of physical media to another or the file structure may change without any need for change in the application programs.

Logical data independence implies that application programs need not be changed if fields are added to an existing records, nor do they have to be changed if fields not used by application programs are deleted. Logical data independence indicated that the conceptual schema can be changed without affecting the existing external schemas.

Data independence in advantageous in the database environment since it allows for change at one level of the database without effecting other levels.

Disadvantage of DBMS:-

A significant disadvantage of the DBMS is cost. In addition to the cost of purchasing or developing the software, the hardware has to be upgraded to allow for extensive programs and the work spaces required for their execution and storage.

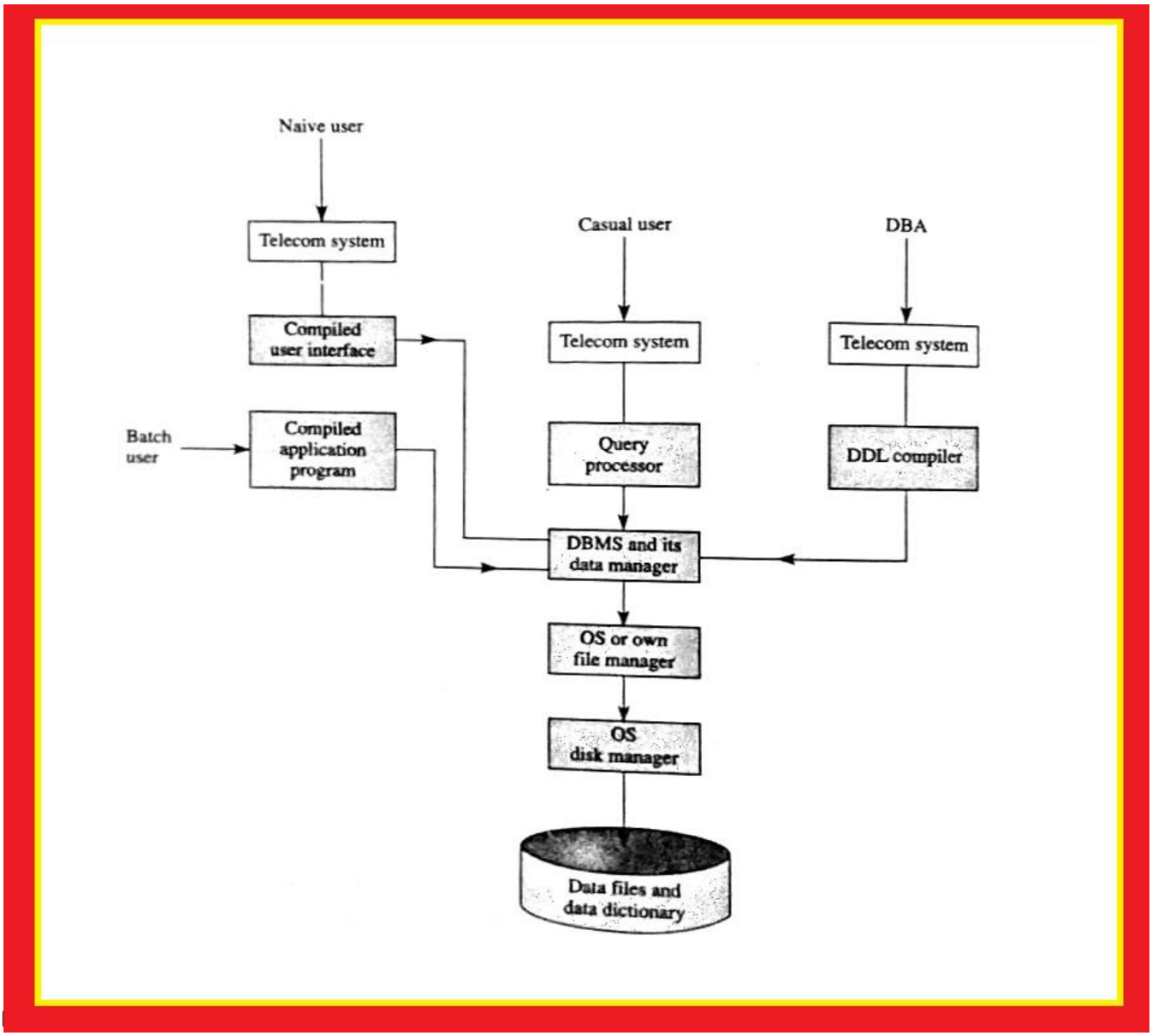
The processing overhead introduced by the DBMS to implement security, integrity and sharing of the data causes a degradation of the response and throughput times.

While centralization reduces duplication, the lack of duplication requires that the database

be adequately backed up so that in the case of failure the data can be recovered. Backup and recovery operations are fairly complicated in the DBMS environment. Ad database system requires a certain amount of controlled redundancies and duplication to enable access to related data items.

Structure of DBMS

For our purposes, we may assume that the database management system is structured and interfaces with various users as shown in following figure. The major components of this system are described below.



Data Definition Language [DDL] Compiler:- The DDL compiler converts the data definition statements into a set of tables. These tables contain the metadata [**Metadata** is "data that provides information about other data". In other words it is data about data.] Concerning the database and are in a form that can be used by other components of the DBMS.

Data Manager :- The data manager is the central software component of the DBMS. It is sometimes referred to as the database control system. One of the functions of the data manager is to convert operations in the user's queries coming directly via the query processor or indirectly via an application program form the user's logical view to a physical file system. The data manager is responsible for interfacing with the file system. In addition, the tasks of enforcing constraints to maintain the consistency and integrity of the data, as well as its security, are also performed by the data manager. It is also entrusted with backup and recovery operations.

File Manager:- Responsibility for the structure of the files and managing the file space rests with the file manager. It is 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.

Disk Manger:- The disk manager is a part of the operating system of the host computer and all physical input and output operations are performed by it.

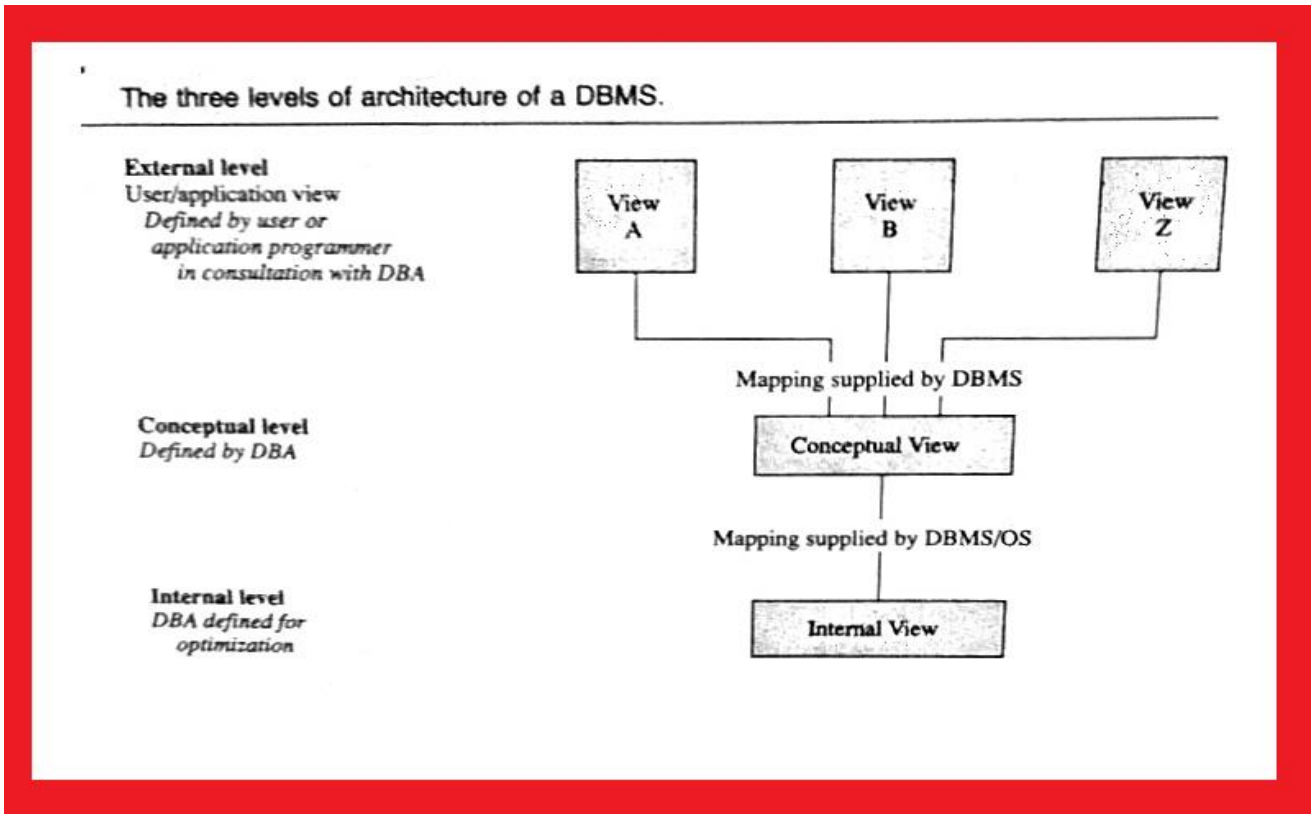
Query Processor:- The query processor is used to interpret the online user's query and convert it into an efficient series of operations in a form capable of being sent to data manager for execution.

The query processor uses the data dictionary [Data dictionary defined in the *IBM Dictionary of Computing*, is a "centralized repository of information about data such as meaning, relationships to other data, origin, usage, and format"] to find the structure of the relevant portion of the database and uses this information in modifying the query and preparing an optimal plan to access the database.

The Three Level Architecture for a DBMS

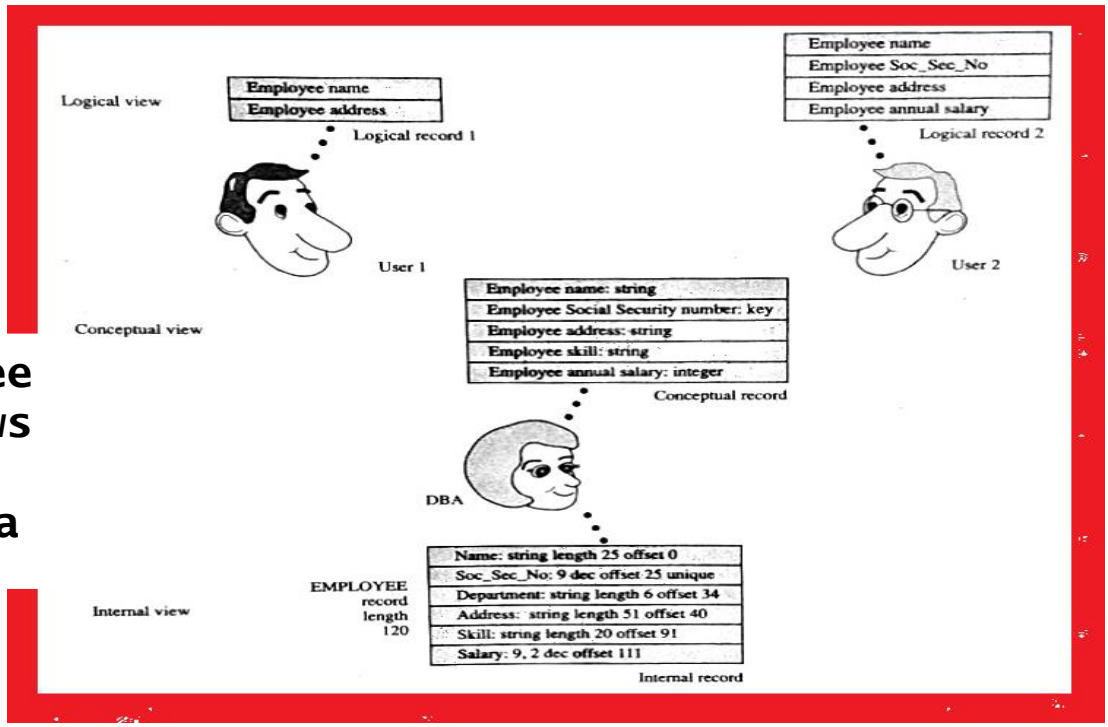
A database management system that provides three levels of data is said to follow three-level architecture. These three levels are the **external level, the conceptual level and the internal level** which are shown in following figures.

The Three Levels of Architecture of a DBMS



Data Models

Three views of data



Hierarchical Model:-

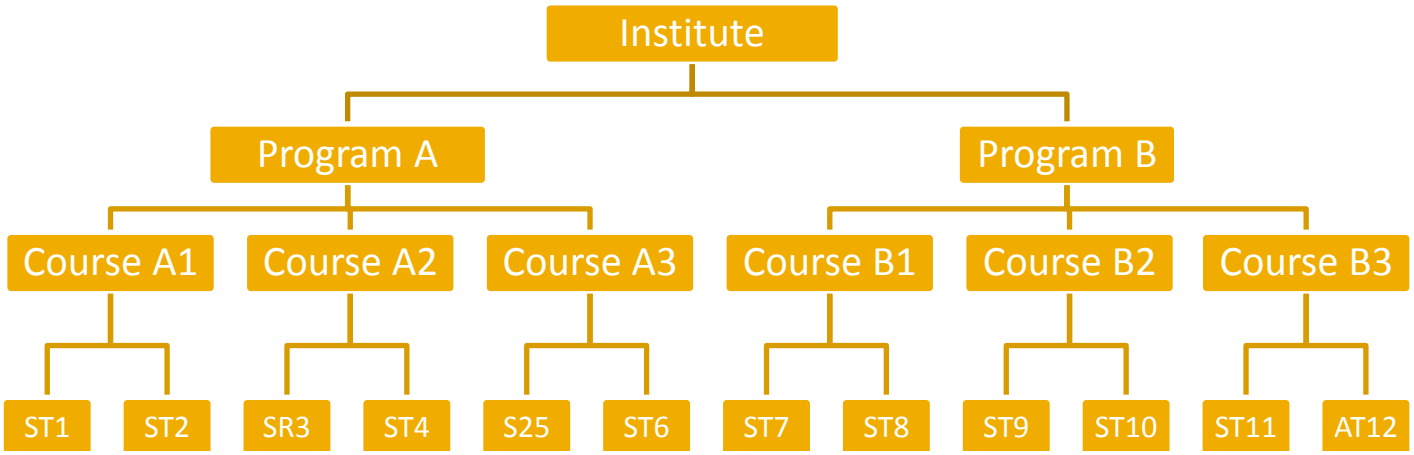
A DMBS belonging to the hierarchical data model uses tree structure to represent relationship among records. A hierarchical consist of a collection of records which are connected with each other through links. Each record is a collection of fields [attributes], each of which contains one data value. A link is an association between two records.

Features of Hierarchical model:-

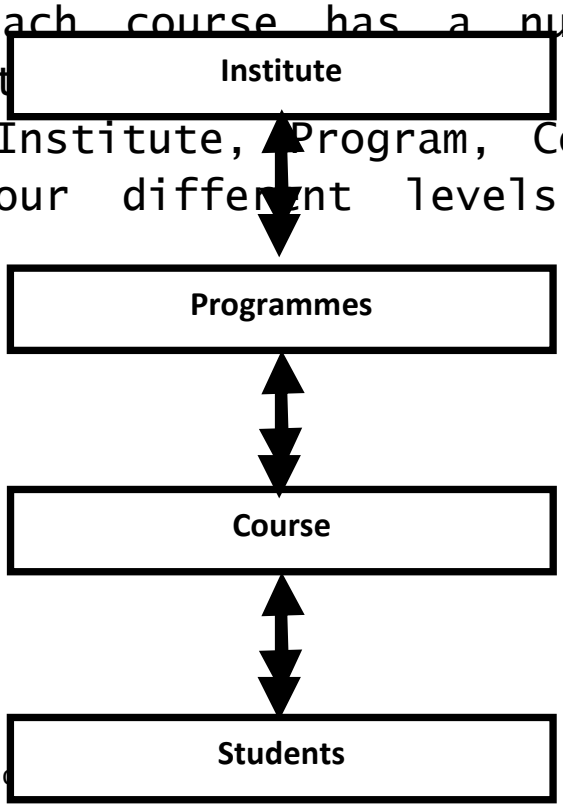
1. Each hierarchical tree can have only one root record type and this record type does not have a parent record type.
2. The root can have any number of child record type and each of which can itself be a root of a hierarchical sub tree.
3. Each child record type can have only one parent record type, thus a M:N relationship cannot be directly expressed between two record types.

4. Data in parent record applies to all its child records.
5. A child record occurrence must have a parent record occurrence, deleting a parent record occurrence requires deleting its entire children record occurrence.

Following figure shows typical database occurrence of a hierarchical structure [tree structure].



The tree structure occur naturally in many data organization because some entities have an intrinsic hierarchical order. For Example, an institute has a number of programs to offer. Each program hs a number of courses. Each course has a number of students registered in it. Figure depicts the four entities types Institute, Program, Course and Student make up the four different levels of hierarchical structute.



Currently hierarchical databases are still widely used especially in applications that require very high performance and availability such as banking and telecommunications. One of the most widely used commercial hierarchical databases is IMS [Information Management System].

NETWORK MODEL:-

The network data model was formalized in the late 1960s by the **D**ata**b**ase **T**ask **G**roup of the **Co**nference on **Da**ta **S**ystem **L**anguage [DBTG/CODASYL]. A network data model is also known as DBTG data model. The DBTG model uses two different structures to represent the database entities and relationships between the entities namely record type and set type.

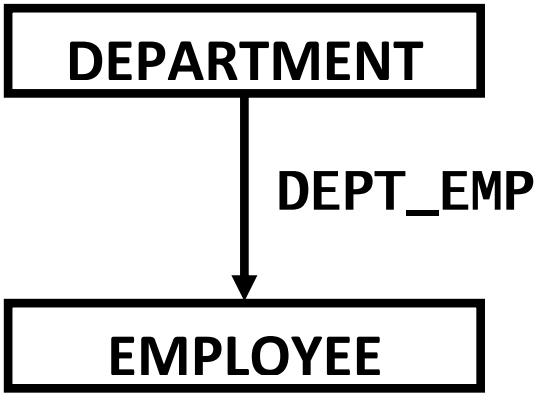
A record type is used to represent an entity type. It is made up of a number of items that represents the attributes of the entity.

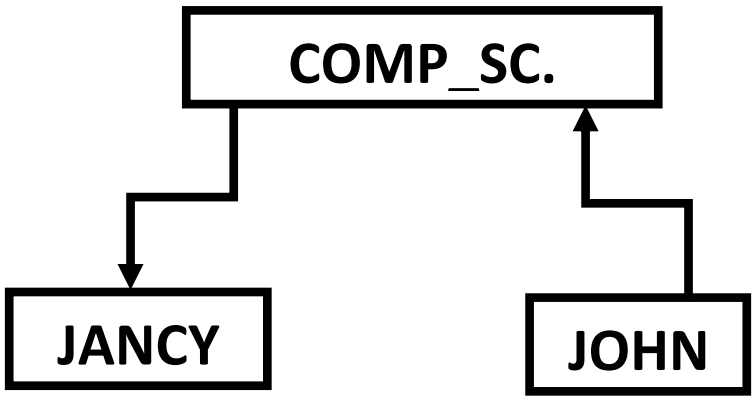
A set type is used to represent a directed relationship between two record types, so called owner record type and member record type.

The set type, like the record type, is named and specifies that there is a one-to-many relationship between the owner and member record types. The set type can have more than one record type as its member, but only one record type is allowed to be the owner in given set type.

Bachman introduced a graphical means called a data structure diagram to denote logical relationship implied by the set here a labeled rectangle represents the corresponding entity or record type. An arrow that connects two labeled rectangles represents a set type. The arrow direction is from owner record type to member record type. Given figure shows two record types DEPARTMETN AND EMPLOYEE and the set type DEPT_EMP with

DEPARTMENT as the owner record type and EMPLOYEE as the member record type.





Relational Model:-

In relational model data is arranged in tables i.e. entities and attributes are expressed in rows and columns indicate the structure, relationship and integrity in the following manner.

1. In any given column of a table, all items are of the same kind.
2. Each item is a simple number or character string.
3. All rows of a table are distinct. In other words, no two rows are identical in every column.
4. Ordering of rows within a table is immaterial.
5. The column of a table are assigned distinct names and the ordering of these column is immaterial.
6. If a tables has N columns, it is said to be of degree N. This is sometimes also referred to as the cardinality of the tables.

Advantage of Relational Approach:-

- Ease of use:- The revision of any information as table consisting of rows and columns quite natural and therefore even first time users find it attractive.
- Flexibility:- Different tables from which information has to be linked and extracted can be easily manipulated by operators such as projection and join to give information in the form in which it is desired.
- Security:- Security control and authorization can also be implemented more easily by moving sensitive attributes in a given table into a separate relation with its own authorization control.
- Data independence:- Data independence is achieved more easily with normalization structure used in a relational database.

Disadvantage of Relational Approach:-

A major disadvantage in the use of relational database is machine performance. If the number of tables between which relationships to be established are large and the tables themselves are voluminous the performance in responding to queries is degraded.