

Chapter 8

Database management System

Concept of Database

A database is a collection of data. A DBMS (**Data Base Management System**) is a set of programs used to create, access and maintain a database. The primary goal of DBMS is to provide an environment that is both convenient and efficient in storing and retrieving database information.

Advantages of DBMS

Database systems are designed to manage large volumes of information. A DBMS consists of a collection of data and a set of programs to access those data. A DBMS has the following advantages

- 1) **Data Redundancy** :- Duplication of data is called data redundancy. In a file management system, data may be repeated in many files. This leads to data redundancy. A DBMS keeps data at one place and all users and applications access the centrally maintained database.
- 2) **Data consistency** :- Data redundancy leads to inconsistency, i.e., when two entries about the same data do not match each other. An inconsistent database provides incorrect data. Inconsistency can be controlled by controlling data redundancy.
- 3) **Efficient data access** :- A DBMS provides an efficient access to database.
- 4) **Data integrity** :- Data integrity refers to correctness of data stored in the database. Data integrity is maintained through implementing rules and procedures. It can also be done by using error checking and validation.
- 5) **Data Security** :- Data security refers to protecting data against accidental loss or disclosure. Data security can be done by using passwords.
- 6) **Sharing of data** :- The data stored in the database can be shared among multiple programs and users.
- 7) **Enforces standard** :- The database enforces standard. These standards may be laid by the organization or individual who uses data.

8)**Recovery**:-A DBMS provides a mechanism for data backup and recovery from hardware failure.

Components of DBMS

A DBMS consists of the following components,

1)**Hardware**

2)**Software**

3)**Data**

4)**Users**

5)**Procedure**

Hardware:-Hardware includes computers(Server) and storage devices for data storage and retrieval.

Software:-The Software consists of application programs and utilities.A DBMS acts as a bridge between the users and database.The users access the database using application programs.

Data:-It is an important component of DBMS.The data in the DBMS is organized in the form of Field,Record and Files.A DBMS provides a centralised control of data.

Users:-The users access the data by using application programs.Depending on the mode of interaction with a DBMS database users are classified into three types

Data Base Administrator(**DBA**),Application Programmer,Sophisticated user and Naive user.

Procedure:- Procedures are rules and instructions that govern the design and use of a database.It may include instruction to start and stop DBMS,backup database,login to database etc.

Structure of DBMS

A database consists of a collection of data.For effective storage and retrieval data is organized as **Field,Record** and **Files**.

Field:-A field is the smallest unit of stored data.For example Roll No,Name,Place etc.

Record:-A record is a collection of related fields.

Files:-A files is a collection of records.

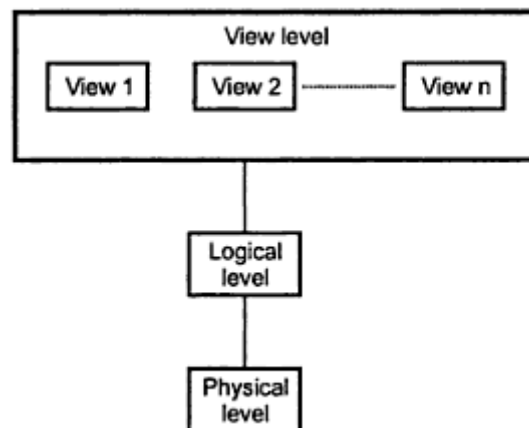
Database Abstraction

The major purpose of database is to provide an abstract view of data.ie, the system hides details of how data is stored and maintained.A database system is designed using three levels of abstraction,Physical Level,Logical Level,View Level.

1)**Physical Level(Internal Level):-**It is the lowest level of abstraction.It describes how data is actually stored in the storage medium.

2)**Logical Level(Conceptual Level):-**Logical level describes what data are stored in the database and the relationship between data.It is also called global view and represents the entire database.It is used by database administrator.

3)**View Level(External Level):-**This is the highest level of database abstraction and is near to the users.It is concerned with the way in which individual users view the data.It describes only a part of entire database.



Three levels of abstraction

Different users in database

Based on the mode of interaction with DBMS the users of a database are classified into four

- Database Administrator(DBA)
- Application Programmer
- Sophisticated Users
- Naive Users

Database Administrator(DBA)

A database administrator is a person who has central control over the database. He is responsible for the installation, configuration, upgrading, administration, monitoring, maintenance, and security of databases in an organization.

Responsibilities of database administrator are

- Database creation.
- Access-method definition.
- Routine maintenance.

Application Programmer:-Application programmers are computer professionals who interact with the database through application programs written in any languages such as C,C++,Java etc.

Sophisticated Users:-Sophisticated users interact with database through queries. They include engineers,anal-yst etc.

Naive users:-Naive users are unsophisticated users .They interact with database by invoking previously written application programs.They are not aware of details of DBMS.

Data Independence

Data independence is the ability to modify the schema definition(Data structure definition) in one level without effecting the schema definition at the next higher level. There are two levels of data independence,Physical data independence and Logical data independence.

Physical data independence

Physical data independence is the ability to modify the physical schema (internal

schema) without effecting conceptual schema.Modification at physical level helps to improve performance.

OR

Physical data independence means the applications need not worry about how the data are physically structured and stored.

Logical data independence

Logical data independence is the ability to modify the conceptual schema without effecting external schema.Logical data independency is difficult to achieve than physical data independency because application programs heavily depend on logical structure of database.

OR

Logical data independence is the ability to modify the conceptual (logical) scheme without causing application programs to be rewritten. Modifications at the logical level are needed whenever logical structure of the database is altered.

Relational data model

A relational model represents database as a collection of relations(**tables**).Each relation has a unique name.A relational model stores data in a tabular form.The main advantage of relational model is that it is simple than other models.

A relational database management system (**RDBMS**) is a database management system based on relational modelA relational database management system (**RDBMS**) is a program that lets you create, update, and administer a relational database.Some of the popular RDBMS are Oracle,MYSQL,DB2 etc.

Terminologies in RDBMS

- 1)**Entity**:-An entity is a real world object such as student,teacher etc.
- 2)**Relation**:-A relation is a collection of data in the form of rows and column.
- 3)**Tuple**:-The row in a table is called tuple.It is also called record.
- 4)**Attribute**:-A column in a table(Relation) is called attribute.
- 5)**Degree**:-The number of columns in a table is called degree.

6)**Cardinality**:-The number of rows in a table is called cardinality.

7)**Domain**:-The set of possible values for a column(**Attribute**) is called domain.

8)**Schema**:-The overall design(**Description**) of a database is called schema.

9)**Instance**:-The collection of data stored in the database at a particular moment is called instance. It is also called **snapshot** or **database state**.

Keys

A key is an attribute or collection of attributes that uniquely identifies each record(Tuple) in a table. A key consisting of one or more attributes is called a composite key(Compound Key).

1)**Candidate Key**:- A candidate key is a column or set of columns that uniquely identifies each record in the table. A table may contain more than one candidate key. For example RollNo + Mark can be considered as a candidate key in the Student table.

2)**Primary Key**:-A primary key is a candidate key which is used to uniquely identify each row in a table. A table can have only one primary key.

3)**Alternate Key**:- An alternate key is a candidate key that is not the primary key.

4)**Foreign Key**:- A foreign key is a field in one table that must match a primary key value in another table. It is used to join two tables together. It is also called reference key.

5)**Super Key**:-A Super key is a set of one or more columns in a table for which no two rows can have the same value. For Example Name and Address can form a Super Key in the Students table.

Relational Algebra

A relational algebra is a collection of operations used to manipulate table content. Relational algebra consists of a set of operations that take one or more relations as input and produce a new relation as output. The fundamental operations in relational algebra are Select, Project, Union, Cartesian Product etc. The select and project are unary operations as they operate on one relation, while other operations are binary operations (**cartesian product, UNION, INTERSECTION and DIFFERENCE**).

SELECT Operation

The SELECT operation selects rows from a table that satisfies a specific condition. It is denoted by $\text{Sigma}(\sigma)$. The select operation gives horizontal subset of a relation.

PROJECT Operation

The PROJECT operation selects attributes (Columns) from a table. It is denoted by $\text{Pi}(\pi)$. The PROJECT operation gives vertical subset of a relation.

Cartesian Product Operation

The cartesian product operation combines tuples from two relations. It is a binary operation and is denoted by \times (cross). It is also called cross product.

UNION Operation

The UNION operation retains a relation consisting of all tuples from both the relations. It is denoted by \cup . The UNION operation takes place between two relations having tables having same number and types of attributes.

R		S	
A	1	A	1
B	2	C	2
D	3	D	3
F	4	E	4
E	5		

R UNION S	
A	1
B	2
C	2
D	3
E	5
F	4
E	4

INTERSECTION Operation

The INTERSECTION operation returns a relation consisting of all tuples common to both the relations. It is a binary operation denoted by \cap .

R		S			
A	1	A	1	A	1
B	2	C	2	D	3
D	3	D	3	E	4
F	4	E	4		
E	5				

R INTERSECTION S**SET DIFFERENCE Operation**

The SET DIFFERENCE operation returns a relation consisting of all tuples appearing in the first relation and not in the second relation. It is denoted by $-$.

R		S		R DIFFERENCE S		S DIFFERENCE R	
A	1	A	1	B	2	C	2
B	2	C	2	F	4	E	4
D	3	D	3	E	5		
F	4	E	4				
E	5						

Conclusion:-

Database administrator is responsible for managing and controlling the activities associated with the database.

The number of tuples in a relation is called cardinality.

Related fields are grouped together to form record.

A foreign key is used to represent relationship between tables.

View level is the highest level of database abstraction.

A database is an organized collection of data.