Relational Database Management Systems – Apr 2016

I. Section-A: **5 X 4 =20 Marks**

1. What is DBMS?

DBMS: Database Management Systems is a collection of inter-related data and set of programs to access those data. The goal of DBMS is to facilitate an interface which is suitable for the users to access and store data.

2. What is Normalization?

Normalization is a process of evaluating and correcting table structures to minimize data redundancies, thereby reducing the likelihood of data anomalies. The normalization process involves assigning attributes to the tables based on the concept of Relational Data Model.

The objective of normalization is to ensure that each table conforms to the concept of well-formed relations. Normalized tables have the following characteristics.

- Each table represents a single subject. For example, a course table will contain only data that directly pertains to courses. Similarly, a student table will contain only student data.

- No data item will be unnecessarily stored in more than one table. The reason for this requirement is to ensure that the data are updated in only one place.

- All non-prime attribute in a table are dependent on the primary key. The reason for this requirement is to ensure that the data are uniquely identifiable by a primary key value.

- Each table is void of insertion, update or deletion anomalies. This is to ensure the integrity and consistency of the data.

3. Compare Table and View

Table: Table are defined in three steps.

1. The name of the table is given.
2. Each column is defined, possibly including column constraints.
3. Table constraints are defined.

Create table Tablename(column-name1, data type(number of characters))
(column-name2, data type(number of characters))
……...
(column-name, data type(number of characters))

View: A definition of a restricted portion of the database. Views are useful for maintaining confidentiality and restricts access to selected parts of the database and for simplifying frequently used
query types.

The formats of create view command is

Create view V As select statement
As = query specification
V= view name

Prashanth Kumar K
(Head-Dept of Computers, IIMC)

Create view stud-view as
(Select stud-no, stud-name
From student
Where percentage<=35)

A view is a database object that represents one or more database tables. It doesn't occupy any table Space.

result set = query (existing database)

4. What is DBA?

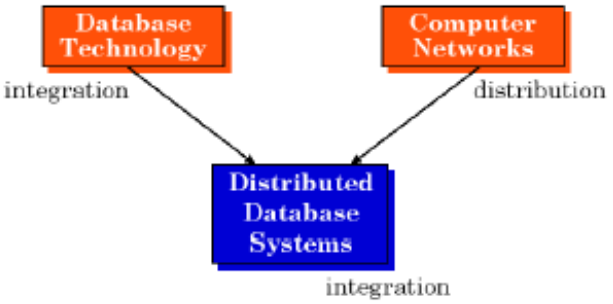
Database Administrator is a person with the responsibility of controlling and protecting the data. The **DBA** should coordinate the design of the database, guide the development and implementation of data security procedures, protect the integrity of data values and make sure system performance is satisfactory.

In a small organization, one person carries out all these responsibilities. Often, these functions are assigned to a group of people. This is most likely in a large organization where DBA responsibilities are divided among several people managed by a chief administrator.

5. What do you mean by Distributed Database?

A distributed database is integrated database which is built on top of a computer network rather than on a single computer. The data which constitute the database are stored at the different sites of the computer network, and the application programs which are run by the computers access data at different sites. Databases may involve different database management systems, running on different architectures, that distributes the execution of transactions.

A Distributed Database Management System (DDBMS) is defined as the software that handles the management of the DDB (Distributed Database) and makes the operation of such a system appear to the user as a centralized database.



6. Write about ORACLE.

The **Oracle RDBMS** is available on many different operating system platforms including **Windows** and **UNIX**. Oracle is a **relational** DBMS - even the data dictionary is simply a collection of tables of data along with indexes and other objects such as sequences and triggers. **SQL** has a basic grammar and syntax. The

functionally of the SQL language is virtually identical across these operating system platforms. Using SQL does not require programming experience, but programming experience can help you conceptualize what a particular SQL command will help to execute and retrieve SQL queries that result data with **SELECT**, **FROM**, **WHERE**, and **ORDER BY commands**. Other command options also exist, but these are the basic ones.

7. What is Foreign Key?

This constraint identifies any column referencing the PRIMARY KEY in another table. It establishes a relationship between two columns in the same table or between different tables. For a column to be defined as a Foreign Key, it should be a defined as a Primary Key in the table which it is referring. One or more columns can be defined as Foreign key.

```
CREATE TABLE product
( product_id number(5) CONSTRAINT pd_id_pk PRIMARY KEY,
product_name char(20),
supplier_name char(20),
unit_price number(10)
);


CREATE TABLE order_items
( order_id number(5) ,
product_id number(5),
product_name char(20),
supplier_name char(20),
unit_price number(10)
CONSTRAINT od_id_pk PRIMARY KEY(order_id),
CONSTRAINT pd_id_fk FOREIGN KEY(product_id) REFERENCES
product(product_id)
);
```

8. What is BCNF?

A relationship is said to be in BCNF if it is already in 3NF and every key is a candidate key. A relation which is in 3NF is almost always in BCNF. This could be same situation when a 3NF relation may not be in BCNF the following conditions are found true.

1. The candidate keys are composite.
2. There are more than one candidate keys in the relation.
3. There are some common attributes in the relation.

II. Section-B: **5 X 10 =50 Marks**

9(a). What are the components of DBMS?

A complete DBMS in an organization consists of the following components.

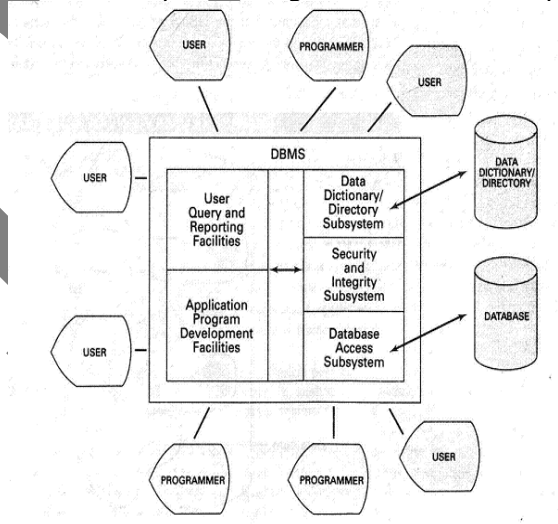
(i). **Hardware:** The hardware is the set of physical devices on which a database resides. It consists of one or more computers, disk drives, CRT terminals, printers, tape drivers, connecting cables, etc.

The computers used for processing the data in the database may be mainframe, mini computers or personal computers. Mainframe and mini computers have traditionally been used on a stand-alone basis to support multiple users accessing a common database. Personal computers are often used with stand-alone databases controlled and accessed by a single user.

Disk drivers are the main storage mechanism for databases. Desktop computers, CRT terminals and printers are used for entering and retrieving information from the database. The success of the database system has been heavily dependent on advances in hardware technology. A very large amount of main memory and disk storage is required to maintain and control the huge quantity of data stored in a database.

(ii). **Software:** A database system includes two types of software:
a. General Purpose database management software usually called the database management system (DBMS).
b. Application software that uses DBMS facilities to manipulate the database to achieve a specific business functions.

Application software is generally written by programmers to solve a specific company problem. It may be written in languages like COBOL or C or it may be written in a language supplied by DBMS like SQL. Application software uses the facilities of the DBMS to access and manipulate data in the database providing reports or documents needed for the information and processing needs of the company.



The DBMS is system software similar to an operating system. It provides a number of services to end users and programmers.

DBMS typically provides most of the following services.

Prashanth Kumar K
(Head-Dept of Computers, IIMC)

1. A central data definition and data control facility known as a data dictionary/directory).
2. Data security and integrity mechanisms.
3. Concurrent data access for multiple users.
4. User-oriented data query, manipulation and reporting capabilities.
5. Programmer-oriented application system development capabilities.

(iii). **Data:** No database system can exist without data. Data can be collected and entered into the Database according to the defined structure.

(iv). **People:** Two different types of people concerned with the database.

They are:

1. Users: Executives, Managers, Staff, Clerical personnel.
2. Practitioners: Database Administrators, Programmers.

(v). **Database:** Database is a collection of inter-related data items that can be processed by one or more application systems.

9(b). Write the Risks and Cost associated with Database.

Database systems have drawbacks.

The following are the Risks & Costs of a database:

(i). **Organizational Conflicts:** Pooling data in a common database may not be politically feasible in some organizations. Certain user groups may not be willing to relinquish control over their data to the extent needed to integrate data. Moreover, the risk involved in data sharing – for example, that one group may damage another group's data – and the potential system problems that may limit a group's access to its own data may be viewed as more troublesome than beneficial. Such people problems could prevent the effectual implementation of a database system.

(ii). **Development Project Failure:** For a variety of reasons, the project to develop a database system may fail. Sometimes management was not fully convinced of the value of the database system in the first place. A database project that seems to be taking too long may be terminated.

A project too large in scope may be almost impossible to complete in a reasonable time. Again, management and users become disenchanted and the project fails.

During the course of a project, key personnel may unexpectedly leave the company. If replacement personnel cannot be found, then the project might not be successfully completed.

(iii). **System Failure:** When the system goes down, all users directly involved in accessing must wait until the system is functional again. This may require a long wait. Moreover, if the system or application software fails, there may be permanent damage to the database. It is very important, therefore to carefully evaluate all software that will have a direct effect on the database to be certain that it is as free as errors as possible. If the organization does not use a database, it is not exposed to this risk, since the data and its software are distributed.

(iv). **Overhead Costs:** The database approach may require an investment in both hardware and software. The hardware to run large DBMS must be efficient and will generally require more main memory and disk storage than simpler file-based system. Tape drivers for rapidly backing up the database are also required. In addition, the DBMS itself may be quite expensive.

The DBMS may also need increase operating costs, since it requires more execution time. For example, an application system using a DBMS will usually execute more slowly than a system not using a DBMS.

(v). **Need for Sophisticated Personnel:** The database administration function requires skilled personnel who are capable of coordinating the needs of different user groups, designing views, integrating those views into a single schema, establishing data recovery procedures and fine tuning the physical structure of the database to meet acceptable performance criteria. There is a risk involved in identification of personnel for the DBA, since if no person having the requisite skills can be found, the DBA function may not be properly performed. This could result in significant problems and may even result in the failure of a database implementation.

10(a). Explain any three Normal Forms.

Normalization is the process of converting a relation into standard form.

It is of different types:

1st Normal Form: A table is in first normal form if all the key attributes have been defined and it contains no repeating groups.

2nd Normal Form: A table is in second normal form (2NF) if and only if it is in 1NF and every non key attribute is fully functionally dependent on the whole of the primary key (i.e. there are no partial dependencies).

3rd Normal Form: A table is in third normal form (3NF) if and only if it is in 2NF and every non key attribute is non transitively dependent on the primary key (i.e. there are no transitive dependencies)

1. Anomalies can occur when a relation contains one or more transitive dependencies.
2. A relation is in 3NF when it is in 2NF and has no transitive dependencies.
3. A relation is in 3NF when 'All non-key attributes are dependent on the key, the whole key and nothing but the key'.

10(b). Write about Relational Algebra.

The Nine relational algebra operators are

1. SELECT: To retrieve specific tuples/rows from a relation.

2. PROJECT: To retrieve specific attributes/columns from a relation.

3. PRODUCT: To obtain all possible combination of tuples from two relations.

4. UNION: To retrieve tuples appearing in either or both the relations participating in the UNION.

5. INTERSECT: To retrieve tuples appearing in both the relations participating in the INTERSECT.

6. DIFFERENCE: To retrieve tuples appearing in the first relation participating in the DIFFERENCE but not the second.

Prashanth Kumar K
(Head-Dept of Computers, IIMC)

7. JOIN: To retrieve combinations of tuples in two relations based on a common field in both the relations.

8. ASSIGNMENT: To store the value in a variable or column.

9. DIVIDE: The result consists of the restrictions of tuples in Table-1 to the attribute names unique to Table-1,i.e., in the header of Table-1 but not in the header of Table-2/

11(a). Write about SQL. Explain any Two DDL and DML Commands.

SQL stands for "Structured Query Language" and can be pronounced as "SQL" or "sequel – (Structured English Query Language)". It is a query language used for accessing and modifying information in the database. IBM first developed SQL in 1970s. Also it is an ANSI/ISO standard. It has become a Standard Universal Language used by most of the relational database management systems (RDBMS). Some of the RDBMS systems are: Oracle, Microsoft SQL server, Sybase etc. Most of these have provided their own implementation thus enhancing it's feature and making it a powerful tool. Few of the SQL commands used in SQL programming are SELECT Statement, UPDATE Statement, INSERT INTO Statement, DELETE Statement, WHERE Clause, ORDER BY Clause, GROUP BY Clause, ORDER Clause, Joins, Views, GROUP Functions, Indexes etc.
In a simple manner, SQL is a non-procedural, English-like language that processes data in groups of records rather than one record at a time. Few functions of SQL are:

- Store data
- Modify data
- Retrieve data
- Modify data
- Delete data
- Create tables and other database objects

DDL Commands:

1. CREATE 2. ALTER 3. DROP 4. TRUNCATE 5.RENAME

DML Commands:

1. INSERT 2. UPDATE 3. DELETE

11(b). Explain about Database Change operations like Arithmetic Operators and Logical Operators

Database Change Operations:

1. INSERT 2. UPDATE 3. DELETE

SQL* ARITHMETIC *Operators:

Operator	Description	Example
+ (unary)	Makes operand positive	SELECT +3 FROM DUAL;
- (unary)	Negates operand	SELECT -4 FROM DUAL;
/	Division (numbers and dates)	SELECT SAL / 10 FROM EMP;

*	Multiplication	SELECT SAL * 5 FROM EMP;
+	Addition (numbers and dates)	SELECT SAL + 200 FROM EMP;
-	Subtraction (numbers and dates)	SELECT SAL - 100 FROM EMP;

SQL LOGICAL OPERATORS

NOT	Returns TRUE if the following condition is FALSE. Returns FALSE if it is TRUE. If it is UNKNOWN, it remains UNKNOWN.	SELECT * FROM EMP WHERE NOT (job IS NULL) SELECT * FROM EMP WHERE NOT (sal BETWEEN 1000 AND 2000)
AND	Returns TRUE if both component conditions are TRUE. Returns FALSE if either is FALSE; otherwise returns UNKNOWN.	SELECT * FROM EMP WHERE job='CLERK' AND deptno=10
OR	Returns TRUE if either component condition is TRUE. Returns FALSE if both are FALSE. Otherwise, returns UNKNOWN.	SELECT * FROM emp WHERE job='CLERK' OR deptno=10

12(a). Write the functions of DBA.

Database Administrator is a person with the responsibility of controlling and protecting the data. The **DBA** should coordinate the design of the database, guide the development and implementation of data security procedures, protect the integrity of data values and make sure system performance is satisfactory.

In a small organization, one person carries out all these responsibilities. Often, these functions are assigned to a group of people. This is most likely in a large organization where DBA responsibilities are divided among several people managed by a chief administrator.

Functions of DBA:

1. **Schema definition**. The DBA creates the original database schema by executing a set of data definition statements in the DDL.

2. **Storage structure and access-method definition**: writing a set of definitions translated by the data storage and definition language compiler

3. **Schema and physical-organization modification**. The DBA carries out changes to the schema and physical organization to reflect the changing needs of the organization, or to alter the physical organization to improve performance.

4. **Granting of authorization for data access**. By granting different types of authorization, the database administrator can regulate which parts of the database various users can access. The authorization information is kept in a special system structure that the database system consults whenever someone attempts to access the data in the system.
5. **Integrity constraint specification:** generating integrity constraints. These are consulted by the database manager module whenever updates occur.

Prashanth Kumar K
(Head-Dept of Computers, IIMC)

12(b). Explain about Database Security.

Database security methods focus on preventing unauthorized users from accessing the database. Because DBMS features that make the database easy to access and manipulate also open doors to intruders, most DBMS include security features that allow only authorized persons or processing that can be accompanied once access is made.

Authentication: - Database access usually requires user authentication and authorization. For user authentication, the first level of security establishes that the person seeking system the user knows, such as log-on number and password, (2) something the user possesses, such as plastic ID card, or (3) a physical representation of the user, such as fingerprint or voiceprint.

Authorization and views: - A view is a means of providing a user with a personalized model of the database. It is also a useful way of limiting a user's access to various positions of the database: Data a user does not need to se are simply hidden from view. This simplifies system usage while promoting security. Executing selects, projections, and joins on existing relations can represent views. The user might also be restricted from seeing any part of the existing relation or from executing joins on certain relations.

Types of Views: - Different types of access authorization may be allowed for a particular view, such as the following:

1. Read authorization: allows reading, but not modification of data.
2. Insert authorization: allows insertion of new data, but no modification of existing data.
3. Update authorization: allows modification of data, but not deletion.
4. Delete authorization: allows deletion of data.

Views and security in SQL: -

CREATE VIEW viewname As (select statement)

Encryption: - The various authentication and authorization measures that are standard for protection access to database may not be adequate for highly sensitive data. In such instances, it may be desirable to encrypt the data. Encrypted data cannot be read by an intruder unless that party knows the method of encryption. Considerable research has been devoted to developing encryption methods.

13(a). Explain DBMS Functions and Capabilities.

The Data Dictionary/Directory: - An effective database system will allow growth and modification in the database without comprising the integrity of its data. The data dictionary/directory (DD/D) aids the accomplishment of this objective by allowing the definitions of data to be maintained separately from the data itself. This allows changes to be made to the data definitions with no effect on the stored data. For example, the subschema used by a particular program could be modified without in any way affecting the stored data. Other benefits provided by the DD/D include these:

- Physical storage structures can be changed without affecting the programs that use the data.
- Passwords and other security measures can be stored in the DD/D to facilitate control over data access.
- Centralized data definition enables easy reporting on the status of the database: Why is responsible for the various data items.
- To yield these benefits, the DD/D usually includes the following features:
- A language for defining entries in the DD/D.
- A manipulation language for adding, deleting, and modifying entries in the DD/D
- Methods for validating entries in the DD/D
- Means for producing reports concerning the data contained in the DD/D.

Data Security and Integrity: -

1) Access Controls: - Access control is an important factor because they are a means of preventing unauthorized access to data. In the data-sharing database environment, good access controls are essential.

2) Concurrency controls: - Concurrency controls are a means of manipulating data integrity in the multi-user environment. Suppose user A and user B both access a given record at (essentially) the same time in order to process a transaction against the record. The DBMS must somehow limit access by one of the users until the others transaction has been completed. Without this type of facility, the accuracy and consistency of the database can rapidly erode.

3) **View Controls: -** It provides an automated means of limiting what a user is allowed to access from a given relation. This is a powerful feature that is commonly provided by relational DBMS. The ease of creating views and the capability of the view facility can be a useful distinguishing factor among DBMSs. The DBMS purchaser may also be interested in whether views can be updated and what limitations may apply.

4) **Encryption:** - It facilitates may be important to institutions whose databases contain very sensitive data. Encryption can also be important for the maintenance of a secure password directory.

5) **Backup and Recovery controls: -**Effective Backup and recovery controls are absolutely essential to efficient operation of the database system. The ease of use of backup and recovery controls, and their completeness, and their reliability should be major factors in the DBMS selection decision.

Query, Data Manipulation, and Reporting Capabilities: -

The DBMS's ability to support reporting requirements, along with users' query and data manipulation needs, is the cornerstone of today's management information systems. A sound DBMS is going to provide the capability to generate structured reports in a variety of formats. In addition, the DBMS will provide a query language that is powerful, yet easy to learn and use. The language should be able to support both planned and unplanned query requirements with short response times.

Support of Specialized Programming Requirements: -

Developing specialized programs to interface with the DBMS requires facilities for supporting program development and program testing. A worthy DBMS will provide a host language for expressing standard procedural program structures or will provide an interface capability for quick prototyping of applications.

Physical Data Organization Options: -

The firm acquiring a DBMS may not wish to involve itself in the details of physical data organization. Instead, it may gauge the efficiency of a DBMS's physical organization by running sample applications.

For those who are interested, however, exploring the physical organization features may be of value. For example, it is known that the inverted list is most efficient in supporting multikey retrieval, whereas the chain list is superior for file updating since there is no need for updating a separate file. Information on other architectural features may be elicited in the process of considering the DBMS's capability to support the types of applications common to the firm.

13(b). Write about Distributed Query Processing.

Distributed Query Processing:

Some database systems support relational databases whose parts are physically separated. Different relations might reside at different sites, multiple copies of a single relation can be distributed among several sites, or one relation might be partitioned into subrelations and these subrelations distributed. In order to evaluate a query posed at a given site, it may be necessary to transfer data between various sites. The key consideration here is that the time required to process such a query will largely be comprised of the time spent transmitting data between sites rather than the time spent on retrieval from secondary storage or computation.

Semijoin: - Suppose the relations R and S shown in Figure. Is stored at sites 1 and 2, respectively. If we wish to respond to a query at site 1 which requires the computation:

JOIN (R, S),

We could transmit all of S from site 2 to site 1 and compute the join at site 1. This would involve the transmission of all 24 values of S.

Site 1		***Site 2***	
R		S	
A1	***A2***	**A2**	**A3**
		A4	
1	3		
1	4	3	13
1	6	16	
2	3	3	14
2	6	16	
3	7	7	13
3	8	17	
3	9	10	14
		16	
		10	15
		17	
		11	15
		16	
		11	15
		16	

Prashanth Kumar K
(Head-Dept of Computers, IIMC)

	12 15 16

Another way would be to compute
T = R [A2]

At site 1; then send T (6 values) to site 2, and compute

U = JOIN (T, S);

And finally send U (9 values) to site 1. We can then compute the desired

JOIN (R, S),

As
JOIN (R, U).

These steps and their results are shown in Figure 12.16. Note that with this approach we have only transmitted 15 values to complete the query.

This example provides a basis for defining a semijoin. The semijoin of R with S is

SEMIJOIN (R, S) = <a projection of those attributes of R that intersect those of S>,

Which is simply that portion of R that joins with S. Therefore,

JOIN (R, S) = JOIN (R, (SEMIJOIN (R, S), S)).

If R and S are at different sites, computing join (R, S) as shown previously saves transmitting data whenever R and S not join completely.