CS8492 DATABASE MANAGEMENT SYSTEMS

L T P C

3 0 0 3

OBJECTIVES

- To learn the fundamentals of data models and to represent a database system using ER diagrams.
- To study SQL and relational database design.
- To understand the internal storage structures using different file and indexing techniques which will help in physical DB design.
- To understand the fundamental concepts of transaction processing- concurrency control techniques and recovery procedures.
- To have an introductory knowledge about the Storage and Query processing Techniques

UNIT I   RELATIONAL DATABASES   10

Purpose of Database System – Views of data – Data Models – Database System Architecture – Introduction to relational databases – Relational Model – Keys – Relational Algebra – SQL fundamentals – Advanced SQL features – Embedded SQL– Dynamic SQL

UNIT II   DATABASE DESIGN   8

Entity-Relationship model – E-R Diagrams – Enhanced-ER Model – ER-to-Relational Mapping – Functional Dependencies – Non-loss Decomposition – First, Second, Third Normal Forms, Dependency Preservation – Boyce/Codd Normal Form – Multi-valued Dependencies and Fourth Normal Form – Join Dependencies and Fifth Normal Form

UNIT III   TRANSACTIONS   9

Transaction Concepts – ACID Properties – Schedules – Serializability – Concurrency Control – Need for Concurrency – Locking Protocols – Two Phase Locking – Deadlock – Transaction Recovery - Save Points – Isolation Levels – SQL Facilities for Concurrency and Recovery.

II Year/IV Semester/2021-2022 DBMS

UNIT IV   IMPLEMENTATIONTECHNIQUES   9

RAID – File Organization – Organization of Records in Files – Indexing and Hashing – Ordered Indices – B+ tree Index Files – B tree Index Files – Static Hashing – Dynamic Hashing – Query Processing Overview – Algorithms for SELECT and JOIN operations – Query optimization using Heuristics and Cost Estimation.

UNIT V   ADVANCEDTOPICS   9

Distributed Databases: Architecture, Data Storage, Transaction Processing – Object-based Databases: Object Database Concepts, Object-Relational features, ODMG Object Model, ODL, OQL - XML Databases: XML Hierarchical Model, DTD, XML Schema, XQuery – Information Retrieval: IR Concepts, Retrieval Models, Queries in IR systems.

TOTAL: 45 PERIODS

OUTCOMES:

Upon completion of the course, the students will be able to:

- Classify the modern and futuristic database applications based on size and complexity
- Map ER model to Relational model to perform database design effectively
- Write queries using normalization criteria and optimize queries
- Compare and contrast various indexing strategies in different database systems
- Appraise how advanced databases differ from traditional databases.

TEXT BOOKS:

1. Abraham Silberschatz, Henry F. Korth, S. Sudharshan, ―Database System Concepts‖,Sixth Edition, Tata McGraw Hill, 2011

2. Ramez Elmasri, Shamkant B. Navathe, ―Fundamentals of Database Systems‖ Sixth Edition, Pearson, 2011.

REFERENCES:

1. C. J. Date, A.Kannan, S. Swamynathan, ―An Introduction to Database Systems‖, Eighth Edition, Pearson Education, 2006.

2. Raghu Ramakrishnan, ―Database Management Systems‖, Fourth Edition, McGraw-Hill College Publications, 2015.

II Year/IV Semester/2021-2022 DBMS

UNIT I

RELATIONAL DATABASES

SYLLABUS: Purpose of Database System – Views of data – Data Models – Database System Architecture – Introduction to relational databases – Relational Model – Keys – Relational Algebra – SQL fundamentals – Advanced SQL features – Embedded SQL– Dynamic SQL

PART A

1. What is database? (R)

A database is a basic electronic storage with collection of interrelated data, organized to provide efficient retrieval. Databases are organized by fields, records and files or tables.

2. What is DBMS? (R)   (NOV 2008)

A database management system (DBMS) is a software package designed to define, manipulate, retrieve and manage data in a database.

3. List out some representative application of databases (U) (MAY 2019)(DEC 2007)

Databases are widely used. Here are some representative applications.

- ➢ Banking
- ➢ Airlines
- ➢ Universities
- ➢ Credit card transactions
- ➢ Telecommunication
- ➢ Finance
- ➢ Sales
- ➢ On-line retailers
- ➢ Manufacturing
- ➢ Human resources

4. What is the purpose of data base system?   (U)   (NOV 2014)

The primary goal of a DBMS is to provide an environment that is both convenient and efficient for people to storing and retrieving information.

II Year/IV Semester/2021-2022 DBMS

5. Define File system. (R)

      Relative data and information is stored collectively in file formats. A file is a sequence of records stored in binary format. A disk drive is formatted into several blocks that can store records. File records are mapped onto those disk blocks.

6. Disadvantage in File Processing System? (R) (DEC 2007)(MAY 2012)(NOV 2016)

- ➢ Data redundancy & inconsistency.
- ➢ Difficult in accessing data.
- ➢ Data isolation
- ➢ Data integrity.
- ➢ Concurrent access is not possible.
- ➢ Security Problems

7. Describe the three levels of data abstraction? (R) (NOV2017)

      Physical schema

          Defines how data is stored
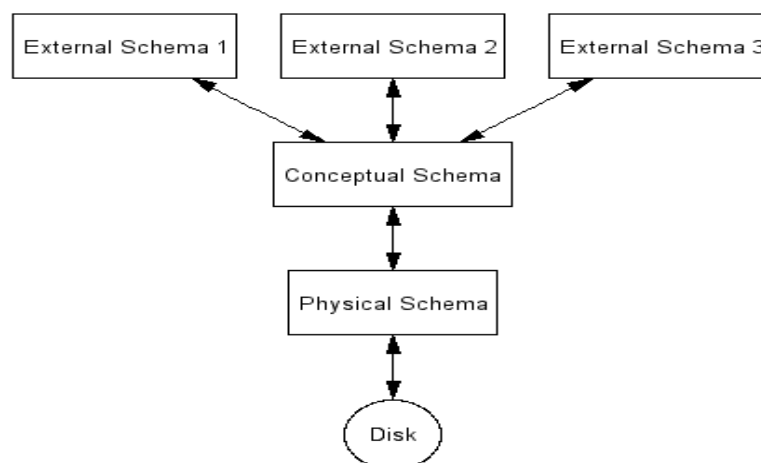
      Conceptual schema or logical schema

          Defines data in terms of a data model

      External schema or view level

          Defines a number of simplified domain-specific views

      DBMS Levels of Abstraction

8. What is a physical level? (R)

The lowest level of abstraction describes how data are stored.

9. What is a logical level? (R)

The next higher level of abstraction, describes what data are stored in database and what relationship among those data.

10. What is a view level? (R)

The highest level of abstraction describes only part of entire database.

11. What is a schema? (R) (MAY 2007)

The structure of a database system, described in a formal language supported by the database management system (DBMS), in a relational database, the schema defines the tables, the fields in each table, and the relationships between fields and tables. Schemas are generally stored in a data dictionary.

12. What are the different types of schemas? (R)

> ➢ Physical Schema
> ➢ Logical Schema
> ➢ External Schema

13. What is an instance? (R)

The collection of information stored in the database at a particular moment is called an instance of the database.

14. What do you mean by data model? (U)(MAY 2011) (MAY 2019)

Data Model can be defined as an integrated collection of concepts for describing and manipulating data, relationships between data, and constraints on the data in an organization.

15. What are the different data models? (R) (MAY 2012) (MAY 2019)

> ➢ Relational Data Model
> ➢ The Entity-Relationship Data Model
> ➢ Object-Based Data Model
> ➢ Semi structured Data Model

16. What is a relational model? (R) (MAY 2010)

The Relational model uses a collection of tables to represent both data and the relationships among those data. Each table has multiple columns, and each column has a unique Value.

17. What is object based data model? (R)

This model is based on collection of objects. An object contains values stored in instance variables with in the object. An object also contains bodies of code that operate on the object. These bodies of code are called methods. Objects that contain same types of values and the same methods are grouped together into classes.

18. What is a semi structured data model? (R)

The semi structured data model permits the specification of data where individual data items of the same type may have different sets of attributes.

19. What is Meta data? (R) (MAY 2008)

Data dictionary contains Metadata-that is, data about data.

20. What is the use of data dictionary? (R)

The DDL gets input as some instructions (statements) and generates some output. The output of the DDL is placed in the data dictionary.

21. Define Relational Database. **(R)**

A relational database is based on the relational model and uses a collection of tables to represent both data and the relationships among those data. It also includes a DML and DDL.

22. What is data base design? (R)

Database design mainly involves the design of the database schema. The design of a complete database application environment that meets the needs of the enterprise being modeled requires attention to a border set of issues.

23. What is physical data independence? (R)

Physical data independence deals with hiding the details of the storage structure from user applications. The application should not be involved with these issues, since there is no difference in the operation carried out against the data.

24. What are different access types in DML? (U)

The types of access are

> ➢ Retrieval of information stored in the database
> ➢ Insertion of new information into the database
> ➢ Deletion of information from the database
> ➢ Modification of information stored in the database

25. What is a procedural DML? (R)

Procedural DMLs require a user to specify what data are needed and how to get those data.

26. What is declarative DML? (R)

Declarative DMLs also referred to as nonprocedural DMLs require a user to specify what data are needed without specifying how to get those data.

28. Define Assertions(R)

An assertion is any condition that the database must always satisfy.

29. What is entity? (R)

It is a 'thing' in the real world with an independent existence.

30. What is relationship set? (R)

The collection (or set) of similar relationships.

31. What is entity set? (R)

It is a collection of all entities of particular entity type in the database.

32. What is week entity? (R) (NOV 2008)(NOV 2016)

An entity set may not have sufficient attributes to form a primary key, and its primary key compromises of its partial key and primary key of its parent entity, then it is said to be Weak Entity set.

33. What is mapping cardinality? (R)

Mapping cardinality express the number of entities to which another entity can be associated via a relationship set.

34. What is functional dependence? (U) (Apr/May2015)

A Functional dependency is denoted by X Y between two sets of attributes X and Y that are subsets of R specifies a constraint on the possible tuple that can form a relation state r of R. The constraint is for any two tuples t1 and t2 in r if t1[X] = t2[X] then they have t1[Y] = t2[Y]. This means the value of X component of a tuple uniquely determines the value of component Y.

35. What is object oriented data model ? (R)

The object –oriented data model is another data model that has seen increasing attention. The object-oriented model can be seen as extending the E-R model with notions of encapsulation, methods and object identity.

36. What is storage manager? (R)

II Year/IV Semester/2021-2022 DBMS

A storage manager is a program module that provides the interface between the low-level data stored in the database and the application programs and queries submitted to the system.

37. What are the components of storage manager ? (U)

> Authorization and integrity manager
> Transaction manager
> File manager
> Buffer manager

38. List out the data structure used to implement the storage manager. (R)

> Data files
> Data dictionary
> Indices

39. What do you mean by functional dependencies? (U) (MAY 2010)(nov2013)(Dec2015)

Let R be a relation variable and let X and Y be arbitrary subsets of the set of attributes of R. then we say that Y is functionally dependent on-in symbols,

$$X \rightarrow Y$$

(read "X functionally determines Y," or simply " X arrow Y ")- if and only if in every possible legal value of R, each X Value has associated with it precisely one Y value.

40. What is a trivial dependency? (R) (MAY 2012)(nov 2013)

A dependency is trivial if it cannot possibly fail to be satisfied.

41. What is a nontrivial dependency? (R) (MAY 2012)

Nontrivial dependencies are the ones that correspond to "genuine" integrity constraints.

42. Define Heath's theorem. (R)

Let R{A,B,C}be a relvar(Relational variable), where A,B, and C are sets of attributes. If R satisfies the FD A$\rightarrow$ B, then R is equal to the join of its projections on {A,B} and {A,C}.

43. Explain steps in query processing? (U)NOV2017

- A given SQL query is translated by the query processor into a low level program called an execution plan
- An execution plan is a program in a functional language
- The physical relational algebra extends the relational algebra with primitives to search through the internal storage structure of DBMS.

Basic Steps in Query Processing

1. Parsing and translation
2. Optimization
3. Evaluation
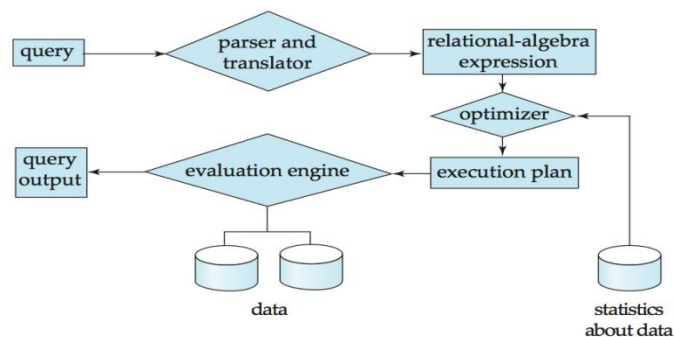
1. Parsing and translation

- Translate the query into its internal form. This is then translated into relational algebra.
- Parser checks syntax, verifies relation.

2. Optimization

- SQL is a very high level language:
  - The users specify what to search for- not how the search is actually done
  - The algorithms are chosen automatically by the DBMS.
- For a given SQL query there may be many possible execution plans.
- Amongst all equivalent plans choose the one with lowest cost.
- Cost is estimated using statistical information from the database catalog.

3. Evaluation

- The query evaluation engine takes a query evaluation plan, executes that plan and returns the answer to that query.

44. What is the use of embedded SQL? (U) (MAY 2012) (NOV 2014)

II Year/IV Semester/2021-2022 DBMS

A fundamental principle underlying embedded SQL, which we call **the dual-mode principle,** is that any SQL statement that can be used interactively can also be embedded in an application program.

45. Discuss about dynamic SQL. (R) (NOV 2014)

Dynamic SQL is a part of embedded SQL. It consists of a set of "dynamic statements"- which themselves are compiled ahead of time- whose purpose is precisely to support the compilation and execution of regular SQL statements that are constructed at run time.

46. What is the difference between static and dynamic SQL? (U) (DEC 2015)(NOV 2016) (NOV2017)

S.No.	Static (embedded) SQL	Dynamic (interactive) SQL
1.	In static SQL how database will be accessed is predetermined in the	In dynamic SQL, how database will be accessed is determined at run time.
2.	It is more swift and efficient.	It is less swift and efficient.
3.	SQL statements are compiled at compile time.	SQL statements are compiled at run time.
4.	Parsing, validation, optimization, and generation of application plan are done at	Parsing, validation, optimization, and generation of application plan are done at
5.	It is generally used for situations where data is distributed uniformly.	It is generally used for situations where data is distributed non-uniformly.
6.	EXECUTE IMMEDIATE, EXECUTE and PREPARE statements are not used.	EXECUTE IMMEDIATE, EXECUTE and PREPARE statements are used.
7.	It is less flexible.	It is more flexible.

47. What is DDL? (R)

The storage structures and access methods used by database system are specified by a set of definition in a special type of DDL called data storage-definition language.

48. What is DML ? (R)

The language, which enable the user to access or manipulate data as organized by appropriate data model.

Procedural DML or Low level: DML requires a user to specify what data are needed and how to get those data.

Non-Procedural DML or High level: DML requires a user to specify what data are needed without specifying how to get those data.

49. What is an Entity? (R)

It is a 'thing' in the real world with an independent existence.

50. What is a query? (NOV2013) (R)

A query is a statement requesting the retrieval of information.

51. Define query language? (R)

A query language is a language in which user requests information from the database.

52. Define Integrity constraint(R)

A constraint (rule) that must remain true for a database to preserve data integrity. Integrity constraints are specified at database creation time and enforced by the database management system.

- ➢ Domain Constraints
- ➢ Referential Integrity
- ➢ Assertions
- ➢ Authorization

53. What are the different component of query processor ? (R)

- ➢ DDL interpreter
- ➢ DML compiler
- ➢ Query evaluation engine

54. State the three levels of abstraction. (R) (MAY 2021)

- **Physical level.** The lowest level of abstraction describes how the data are actually stored.
- **Logical level.** The next-higher level of abstraction describes what data are stored in the database, and what relationships exist among those data.
- **View level.** The highest level of abstraction describes only part of the entire database.

55. What are the various types of keys in the database? (R) (MAY 2021)

- Super keys

- Candidate keys

- Primary keys

- Composite Key

- Secondary key

- Foreign key

- Simple key

PART B

1. Explain the system structure of a database system with neat block diagram. **(R)**

   (DEC 2007), (MAY 2010) (MAY 2012)

2. What is a data model? Explain various data model for describing the design of a **(U)**

   Database at the logical level**. (APRIL 2008), (MAY 2010)**

3. Explain the different between physical and logical data independence with an example. **(U)**

   (APRIL 2008)

4. Explain the characteristics of database approach. **(U) (APRIL 2008)**

5. Describe the three-schema architecture. Why do we need mapping between schema

   Levels? How do different schema definition languages support this architecture?**(R)**
(Dec2008)

6. What is the notation used in E-R diagram? Explain the E-R model structure with Example.
(U) (NOV 2014)

7. Explain the role and functions of the database administrator. **(R)**

8. Explain the following **(R) (APRIL 2008) (NOV 2014)**

   i. Data Definition Language.

   ii. Data Manipulation Language.

   iii. Data control Language.

   iv. Views

II Year/IV Semester/2021-2022 DBMS

9.   Develop an Entity Relationship model for a library management system. Clearly
     State the problem Definition, Description, Business Rule and any assumption you
     make**. (AP)(MAY 2009) (NOV 2014)**

10.   Construct an ER diagram for a car insurance company, banking that has a set of customers,
      each of owns one/more care. Each car has associated with it zero to any number of
      recorded accidents.**(AP) (MAY 2012) (NOV 2016)**

11.   Discuss the various disadvantages in the file system and explain how it can be overcome by
      the database system**. (AN) (MAY 2010)**

12. Construct an ER diagram for an employee payroll system. **(AP) (MAY 2010) (NOV 2014)**

13. Define Normalization. Describe the procedure of Normalization considering all the
 forms with suitable illustrations. **(AP) (MAY 2007)**

14. Give the comparison between BCNF and 3 NF**. (U) (DEC 2007) , (MAY 2010)**

15. What is FD (Functional Dependencies)? Explain the role of FD in the process of
 Normalization. **(U) (MAY 2007)**

16.What is normalization? Explain 1NF, 2NF, 3NF and BCNF with simple example**.**

(U) (MAY 2007), (MAY 2010) (MAY 2012)(NOV 2016)

17. Justify the need for normalization with example. **(U) (MAY 2007)**

18. Describe about the Multi valued dependencies and fourth normal form with suitable example.

(U)(MAY 2012)

19.State the need for Normalization of a Database and Explain the various Normal Forms(1st
,2nd,3rd BCNF,4th,5th and Domain Key)with suitable examples**.(U)(Dec2015)**

20.Draw E-R diagram for the "Restaurant Menu Ordering System", which will facilitate the food
items ordering and services within a restaurant. The entire restaurant scenario is details as
follows. The customer is able to view the food items menu, call the water, place orders and
obtain the final bill through the computer kept in their table. The Waiters through their wireless
tablet PC are able to initialize a table for customers, control the table functions to assist
customers, orders, send orders to food preparation staff (chef) and finalize the customer's bill.
The Food preparation staffs (Chefs), with their touch-display interfaces to the system, are able to

view orders sent to the kitchen by waiters. During preparation, they are able to let the waiter know the status of each item, and can send notifications when items are completed. The system should have full accountability and logging facilities, and should support Supervisor actions to account for exceptional circumstances, such as a meal being refunded or walked out on. **(AP)(Dec2015)**

21. Explain the basic relational Algebra operations with the symbol used and an example for each**.(R)(DEC 2008) (MAY 2012)**

22. Explain Select , project and Cartesian product in relational algebra with an examples **(U)(NOV 2016)**

23. Differentiate between foreign key constraints and referential integrity constraints with suitable example. ii) Distinguish between lossless-join decomposition and dependency preserving decomposition. **(AN) NOV2017**

24. State and explain the architecture of DBMS. Draw the ER diagram for banking systems. (Home loan applications). **(AN)NOV2017**

25. Explain the three different groups of data models with suitable examples. **(R) (MAY 2019)**

26. Describe about the static and dynamic SQL in detail. **(R) (MAY 2019)**

27. State and explain the architecture of DBMS. Discuss about the people who deals with database. **(U) (MAY 2021)**

28. What are the several parts of SQL query language? What are the basic built in types used during SQL create statement? State and give example for basic structure of SQL queries? **(U) (MAY 2021)**

COURSE OUTCOMES

 Design Databases for applications and use the Relational model, ER diagrams.

UNIT II

DATABASE DESIGN

SYLLABUS: Entity-Relationship model – E-R Diagrams – Enhanced-ER Model – ER-to-Relational Mapping – Functional Dependencies – Non-loss Decomposition – First, Second, Third Normal Forms, Dependency Preservation – Boyce/Codd Normal Form – Multi-valued Dependencies and Fourth Normal Form – Join Dependencies and Fifth Normal Form

PART A

1. **What are the aspects of relational model?(R)**

    The Three aspects of relational model are:

    - ❖ Structural aspect: The data in the database is perceived by the user as tables, and nothing but tables.
    - ❖ Integrity aspect: Those tables satisfy certain integrity constraints.
    - ❖ Manipulative aspect: The operators available to the user for manipulating those tables- for example, for purposes of data retrieval – are operators that derive tables from tables. Of those operators, three particularly important ones are select, project and join.

2. **Mention the important points about relational databases. (R)**

    A set of important points about relational databases:

    a. Relational databases store data in the form of tables (logically).
    b. The rows of a table are called as tuples.
    c. The columns of a table are known as attributes.
    d. Every attribute has a data type associated with it.
    e. Every attribute has a domain which provides the set of all possible values that can be stored as values for that attribute.
    f. Tables are called as relations.
    g. The table names are called as relational variable

3. **What is an entity-relationship model? (R) (MAY 2008)**

    This data model is based on real world that consists of basic objects called entities and of relationship among these objects. Entities are described in a database by a set of attributes.

4. **What is object based data model? (R)**

This model is based on collection of objects. An object contains values stored in instance variables with in the object. An object also contains bodies of code that operate on the object. These bodies of code are called methods. Objects that contain same types of values and the same methods are grouped together into classes.

5. What is a polymorphic operator? (R)

An operator is said to be polymorphic if it is defined in terms of some parameter P and the arguments corresponding to P can be of different types on different invocations.

6. What is a candidate key? (R)

Let K be a set of attributes of relvar R. Then K is a candidate key for R if and only if it has both of the following properties: Uniqueness and Irreducibility

7. What is a primary key? (R)

It is possible for a given relation to have two or more candidate keys. In such a case, the relational model has historically required that exactly one of those keys be chosen as the primary key.

8. What is an alternate key? (R)

It is possible for a given relation to have two or more candidate keys. In such a case, the relational model has historically required that exactly one of those keys be chosen as the primary key and the others are then called alternate keys.

9. What is a foreign key? (R)

A foreign key is a set of attributes of some relvar R2 whose values are required to match values of some candidate key of some relvar R1.

10. What is a trigger? (R)

Triggered procedures are precompiled procedures that are stored along with the database and invoked automatically whenever some specified event occurs.

11. What are the relational operators? (R)

Union, Intersect, Difference, Product, Select, Project, Join, Divide are the relational operators.

12. Define Cartesian product. (R)

II Year/IV Semester/2021-2022 DBMS

Cartesian product of two relations a and b, a TIMES b, where a and b have no common attribute names, to be a relation with a heading that is the union of the headings of a and b and with a body appearing in a and a tuple appearing in b.

13. What is a range variable? (R)

A range variable is a variable that "ranges over" some specified relation. If range variable V ranges over relation r, then, at any given time, the expression "V" denotes some tuple of r.

14. Discuss about query-by-example? (U)

Query-By-Example syntax, which is attractive and intuitively very simple, is based on the idea of making entries in blank tables.

15. What do you mean by type constraints? (R)

A type constraint is, precisely, a definition of the set of values that constitute a given type.

16. What do you mean by database constraints? (R)

A database constraint is a constraint on the values a given database is permitted to assume.

17. What do you mean by relation constraints? (R)

A relation constraint is a constraint on the values a given relvar is permitted to assume.

18. What do you mean by attribute constraints? (R)

An attribute constraint is a constraint on the values a given attribute is permitted to assume.

19. What is referential integrity? (R)

Referential integrity database must not contain any unmatched foreign key values.

20. What is a transition constraint? (R)

A transition constraint is a constraint on the legal transitions that a given variable-in particular, a given relvar or a given database-can make from one value to another.

21. Give the usage of the rename operation with an example.(MAY 2010) (U)

RA _S1.Name;S2.Name(_S1.Address = S2.Address(_S1(Students) _ _S2(Students))).

22. What do you mean by weak entity set? .(MAY 2010) (R)

II Year/IV Semester/2021-2022 DBMS

A **weak entity** is an entity that cannot be uniquely identified by its attributes alone; therefore, it must use a foreign key in conjunction with its attributes to create a primary key. The foreign key is typically a primary key of an entity it is related to.

23. What is the difference between Data Integrity and Data Security?(nov 2013) (U)

Data integrity and data security are two different aspects that make sure the usability of data is preserved all the time. Main difference between integrity and security is that integrity deals with the validity of data, while security deals with protection of data. Backing up, designing suitable user interfaces and error detection/correction in data are some of the means to preserve integrity, while authentication/authorization, encryptions and masking are some of the popular means of data security. Suitable control mechanisms can be used for both security and integrity.

24. Which operators are called as unary operators and why are they called so(nov 2013)(U)

An operator that takes a single operand in an expression or a statement. The unary operators in C# are +, -,!, ~, ++, -- and the cast operator.

25. What is atomicity ? (R)

Several operations on the database form a single logical unit of work is called atomicity.

26. What is a transaction ? (R)

A transaction is a collection of operations that performs a single logical function in a database application.

27. What is DBA ? (R) (MAY 2010),(MAY 2011)

A person who has such central control over the system is called a database administrator (DBA).

28. What are the responsibilities of DBA? (U) (MAY 2011)

➤ Schema definition
➤ Storage structure and access-method definition
➤ Schema and physical organization modification
➤ Granting of authorization for data access
➤ Routine maintenance

29. Define Query processor. (R)

A relational database consists of many parts, but at its heart are two major components: the storage engine and the query processor.

II Year/IV Semester/2021-2022 DBMS

30. What is the difference between static and dynamic SQL? (U) (DEC 2015)(NOV 2016) (NOV2017)

S.No.	Static (embedded) SQL	Dynamic (interactive) SQL
1.	In static SQL how database will be accessed is predetermined in the embedded SQL statement.	In dynamic SQL, how database will be accessed is determined at run time.
2.	It is more swift and efficient.	It is less swift and efficient.
3.	SQL statements are compiled at compile time.	SQL statements are compiled at run time.
4.	Parsing, validation, optimization, and generation of application plan are done at compile time.	Parsing, validation, optimization, and generation of application plan are done at run time.
5.	It is generally used for situations where data is distributed uniformly.	It is generally used for situations where data is distributed non-uniformly.
6.	EXECUTE IMMEDIATE, EXECUTE and PREPARE statements are not used.	EXECUTE IMMEDIATE, EXECUTE and PREPARE statements are used.
7.	It is less flexible.	It is more flexible.

31. What are the difference between DDL, DML and DCL commands? (U) (NOV 2015)

Data Definition Language (DDL) statements are used to define the database structure or schema. Some examples:

    CREATE - to create objects in the database
    ALTER - alters the structure of the database
    DROP - delete objects from the database
    TRUNCATE - remove all records from a table, including all spaces allocated for the records are removed
    COMMENT - add comments to the data dictionary
    RENAME - rename an object

Data Manipulation Language (DML) statements are used for managing data within schema objects. Some examples:
    SELECT - retrieve data from the a database
    INSERT - insert data into a table
    UPDATE - updates existing data within a table
    DELETE - deletes all records from a table, the space for the records remain

II Year/IV Semester/2021-2022 DBMS

MERGE - UPSERT operation (insert or update)
CALL - call a PL/SQL or Java subprogram
EXPLAIN PLAN - explain access path to data
LOCK TABLE - control concurrency

Data Control Language (DCL) statements. Some examples:

GRANT - gives user's access privileges to database
REVOKE - withdraw access privileges given with the GRANT command

Transaction Control (TCL) statements are used to manage the changes made by DML statements. It allows statements to be grouped together into logical transactions.
COMMIT - save work done
SAVEPOINT - identify a point in a transaction to which you can later roll back
ROLLBACK - restore database to original since the last COMMIT
SET TRANSACTION - Change transaction options like isolation level and what rollback segment to use.

32.State the Need for Query Optimization. (R)

- Process of selecting an efficient execution plan for evaluating the query.

- refers to the process of finding lowest cost method of evaluating a given query.

- Need for being logically consistent because the least cost plan will always be consistently low.

33. What is Data Definition Language Give example (R) (NOV 2016)

Data Definition Language (DDL) statements are used to define the database structure or schema. Some examples:
CREATE - to create objects in the database
ALTER - alters the structure of the database
DROP - delete objects from the database
TRUNCATE - remove all records from a table, including all spaces allocated for the records are removed
COMMENT - add comments to the data dictionary
RENAME - rename an object

34. Define first normal form.(R)

A relvar(Relational variable) is in 1NF if and only if, in every legal value of that relvar(Relational variable), every tuple contains exactly one value for each attribute.

35. Define second normal form.(R)

A relvar(Relational variable) is in 2NF if and only if it is in 1NF and every nonkey attribute is irreducibly dependent on the primary key.

II Year/IV Semester/2021-2022 DBMS

36. Define third normal form. (R)

A relvar(Relational variable) is in 3NF if and only if it is in 2NF and every nonkey attribute is nontransitively dependent on the primary key.

37. What is dependency preservation? (R)

Normalization procedure should decompose relvars(Relational variable) into projections that are independent in Rissanen's sense has come to be known as dependency preservation.

38. Define Boyce/Codd normal form. (R) (MAY 2012)

A relvar(Relational variable) is in BCNF if and only if every nontrivial, left-irreducible FD has a candidate key as its determinant (OR) A relvar(Relational variable) is in BCNF if and only if every determinant is a candidate key.

39. What is multi-valued dependence? (R)

Let R be a relvar(Relational variable), and let A, B and C be subsets of the attributes of R. Then we say that B is multi-dependent on A-in symbols, A$\rightarrow\rightarrow$B

40. Define Fagin theorem. (R)

Let R{A,B,C} be a relvar(Relational variable), where A, B, and C are sets of attributes. Then R is equal to the join of its projections on {A, B} and {A, C} if and only if R satisfies the MVDs A$\rightarrow\rightarrow$B|C.

41. Define fourth normal form. (R)

Relvar(Relational variable) R is in 4NF if and only if, whenever there exist subsets A and B of the attributes of R such that the nontrivial MVD A$\rightarrow\rightarrow$B is satisfied, then all attributes of "R" are also functionally dependent on A.

42. What is a join dependency? (R)

Let R be a relvar(Relational variable), and let A, B,…. .Z be subsets of the attributes of R. Then we say that R satisfies the JD

*{A, B …Z}

if and only if every legal value of R is equal to the join of its projections on A, B,……Z.

43. Define fifth normal form. (R)

A relvar(Relational variable) R is in 5NF-also called projection-join normal form(PJ/NF)-if and only if every nontrivial join dependency that is satisfied by R is implied by the candidate key(s) of R, where

a. The join dependency *{A, B,…….Z} on R is trivial if and only if at least one of   A,

   B,…Z is the set of all attributes of R.

b. The join dependency *{A, B,…….Z} on R is implied by the candidate key(s)

   of R if and only if each of A,B,…..,Z is a super key of R.

44. What is de-normalization? (R)

Let R1, R2,….., Rn be a set of relvars(Relational variable). Then de-normalizing those relvars means replacing them by their join R, such that for all i (i=1, 2,….., n) projecting R over the attributes of Ri is guaranteed to yield Ri again.

45. What is normalization? (MAY 2010) (R)

Normalization is the process of structuring relational database schema such that most ambiguity is removed.

46. Give the Properties of Decomposition. (R) (May 2019)

Following are the properties of Decomposition,

1.LosslessDecomposition

2.DependencyPreservation

3. Lack of Data Redundancy

47. Define the term entity set and relationship set? (R) (May 2019)

Relationship set

The collection (or set) of similar relationships.

Entity set

It is a collection of all entities of particular entity type in the database.

48.Define functional dependency? (R) (May 2021)

The functional dependency is a relationship that exists between two attributes.

It typically exists between the primary key and non-key attribute within a table.

X → Y

The left side of FD is known as a determinant, the right side of the production is known as a dependent.

49. What is ER diagram? (R) (May 2021)

- E-R diagram is the short form of "Entity-Relationship" diagram.

- An E-R diagram efficiently shows the relationships between various entities stored in a database.

- E-R diagrams are used to model real-world objects like a person, a car, a company etc. and the relation between these real-world objects.

PART B

1.Discuss about Tuple Relational Calculus and Domain Relational Calculus. **(R) (DEC 2008)** **(MAY 2012)**

2.What are aggregate functions? Explain five built-in aggregate functions. **(U) (MAY 2008)**

3.Consider the following relations for a company database Application**.(AP)(MAY 2009)**

Employee (**Eno**, Name, Sex, DOB, Doj, Designation, Basic_Pay, Dept_No)
Department (**DeptNo**,Name)
Project (**ProjNo**, Name, Dept_No)
Works for (**Eno,ProjNo,Date,** Hours)
The attributes specified for each relation is self-explanatory. However the business rules are stated as follows. A department can control a project. An employee can work on any number of projects on a day. However an employee cannot work more than once on a project he/she worked on that day. The primary key are underlined.
(i). Identify the forging keys, Develop DDL to implement the above schema.

(ii) Develop an SQL query to list the department number and the number of Employee in each department.

(iii) Develop a View that will keep track of the department number, the number of employees in the department, and the total basic pay expenditure for each department.

(iv) Develop an SQL query to list the details of employees who have worked in more than three projects on a day.

4. Explain how dangling tuple may arise and also explain problems that they Cause**. (U)**

(MAY 2008)

5. Write detailed notes on the following: **(U) (APRIL 2008) (MAY 2010)**

      i. Distributed Databases.

      ii. Client/server database.

6. Explain the architecture of Distributed Database Management System**. (R) (DEC 2007)**

7. Create an EMPLOYEE table and write the steps for various functions in an SQL Like add, update, delete, save and join the various attributes of an EMPLOYEE Table**.**

(C)(MAY 2007)

8. Explain the use of trigger with your own example. **(U) (MAY 2010)**

9. What is a view? How can it be created? Explain with an example. **(R) (MAY 2010)**

10. Briefly present a survey on Integrity and Security. **(U) (MAY 2012)**

11. Discuss about query processing and query optimization**. (U) (NOV 2014)**

12.Consider a student registration database comprising of the given below table schema**.(AP)(Nov/Dec2015)**

Student File

Student number	Student name	Address	Telephone

Course File

Course Number	Description	Hours	Professor Number

Professor File

Professor Number	Name	Office

Registration File

Student Number	Course Number	Date

Consider a suitable example of tuples/records for the above mentioned tables and write DML statements (SQL) to answer for the queries listed below:

(i) Which courses does a specific professor teach?

(ii)What courses are taught by two specific professors?

(iii) What teaches a specific course and where is his/her office?

(iv) For a specific student number, in which courses is the student registered and what is his/her name?

(v) Who are he professors for a specific student?

(vi) Who are he students registered in a specific course?

12. Discuss about the Join order optimization and Heuristic optimation algorithms**. (U) (Nov/Dec2015)**

II Year/IV Semester/2021-2022 DBMS

13. Explain Query optimization with an example **(U)(NOV 2016)**

14. Let relations r1(A,B,C)and r2(C,D,E) have the following properties r1 has 20000 tuples r2 has 45000 tupes ,25 tuples of r1, fit on one block and 30 tuples of r2 fit on one block .Estimate the number of block transfers and seeks required,using each of the following join strategies for

r1∞r2:

(i)Nested loop join

(ii)Block Nested loop join

(iii) Merge join

(iv) Hash join   **(AP)(NOV 2016)**

15. What is Embedded SQL? Give example**(R)(Nov 2016)**

16. a) i) State and explain the command DDL, DML, DCL with suitable example. ii) Justify the need of embedded SQL. Consider the relation student (student no, name, mark and grade). Write embedded dynamic SQL statements in C language to retrieve all the students' records whose mark is more than 90. (AN)**NOV2017**

17. Explain the catalog information for cost estimation for selection and sorting operation in data. (R)**NOV2017**

18. What is Normalization? Explain in detail about all normal forms. **(R) (May – 2019)**

19. Briefly discuss about the functional dependency concepts. **(R) (May – 2019)**

20. Explain the following terms briefly: attribute, domain, entity relationship, entity set, relationship set, one-to-many relationship, many-to-many relationship, participation constraint, overlap constraint, covering constraint, weak entity set, aggregation and role indicator. **(R) (May – 2021)**

21. Consider the following relations: **(AP) (May – 2021)**

Sailors (sid:integer, sname:string, rating:integer,

age:real)Boats (*bid:*integer, *bname*:string,

color:string) Reserves(*sid*:integer, *bid*:integer,

day:date)

Write the SQL statement for the following queries :

 i) Find all sailors with a rating above 7.
 ii) Find the sids of sailors who have reserved a red boat.
 iii) Find the colors of boats reserved by lubber.

II Year/IV Semester/2021-2022 DBMS

iv) Find the names of sailors who have reserved at least one boat.

COURSE OUTCOMES

Construct DDL, DML, TCL and DCL commands for different databases with various constraint

UNIT III

TRANSACTIONS

SYLLABUS:**Transaction Concepts – ACID Properties – Schedules – Serializability – Concurrency Control – Need for Concurrency – Locking Protocols – Two Phase Locking – Deadlock – Transaction Recovery - Save Points – Isolation Levels – SQL Facilities for Concurrency and Recovery.**

PART – A

1. What is a catalog? (R)

    The catalog contains detailed information, sometimes called descriptor information or Metadata, regarding the various objects that are of interest to the system itself.

2. Define a transaction. (R)

    A transaction is a logical unit of work, typically involving several database operations. A transaction begins when **BEGIN TRANSACTION** is executed and terminates when **COMMIT** or **ROLLBACK** is executed.

    3. What is the use of embedded SQL? (U) (MAY 2012) (NOV 2014)

    A fundamental principle underlying embedded SQL, which we call **the dual-mode principle,** is that any SQL statement that can be used interactively can also be embedded in an application program.

4. What is a cursor? (R)

II Year/IV Semester/2021-2022 DBMS

A cursor consists essentially of a kind of (logical) pointer - a pointer in the application, that is, not one in the database - that can be used to run through a collection of rows, pointing to each of the rows in turn and thereby providing addressability to those rows one at a time.

5. **Write short note on OPEN, FETCH and CLOSE statements. (R)**

   OPEN:

   EXEC SQL OPEN < Cursor name>;

   Opens the specified cursor. A set of rows is thus identified and becomes the current active set for the cursor.

   FETCH:

   EXEC SQL FETCH <Cursor name>

   INTO <host variable reference commalist>;

   Advances the specified cursor to the next row in the active set.

   CLOSE:

   EXEC SQL CLOSE <Cursor name>;

   Closes the specified cursor. The cursor now has no current active set.

6. **Define recovery in a database system.(R)**

   Recovery in a database system means, primarily, recovering the database itself; that is, restoring the database to a correct state after some failure has rendered the current state incorrect, or at least suspect.

7. **What is a transaction? (MAY 2010) (R)**

   A transaction is a logical unit of work; it begins with the execution of a BEGIN TRANSACTION operation, and ends with the execution of a COMMIT or ROLLBACK operation.

8. **What is a commit point? (R)**

   A commit point corresponds to the end of a logical unit of work, and hence to a point at which the database is supposed to be in a correct state.

9. **Discuss the ACID properties. (R) (MAY 2010) (NOV 2014)(Dec2015) (NOV 2017)**

   ACID properties are

   (i)   Atomicity: Transactions are atomic

     (ii)   **Correctness:** Transactions transform a correct state of the database into another correct state, without necessarily preserving correctness at all intermediate points

     (iii)   **Isolation:** Transactions are isolated from one another. That is, even though in general there will be many transactions running concurrently, any given transaction's updates all connected from all the rest, until that transaction commits. Another way of saying the same thing is that, for any two distinct transactions A and B , A might see B's updates or B might see A's updates, but not both.

     (iv)   **Durability:** Once a transaction commits, its updates persist in the database, even if there is a subsequent system crash.

10. What is a system failure? (R)

System failures (e.g., power outage), which affect all transactions currently in progress but do not physically damage the database. A system failure is sometimes called a soft crash.

11. What is a media failure? (R)

Media failures (e.g. head crash on the disk), which do cause damage to the database or some portion thereof, and affect at least those transactions currently using that portion. A media failure is sometimes called a hard crash.

12. What is a check point record? (R)

The checkpoint record contains a list of all transactions that were in progress at the time the checkpoint was taken.

13.   Expand ARIES. (R)

The name ARIES stands for "Algorithms for Recovery and Isolation Exploiting Semantics."

14.   What are the three broad phases of ARIES? (U)

ARIES operates in three broad phases:

➢ **Analysis:** Build the REDO and UNDO lists
➢ **Redo:** Start from a position in the log determined in the analysis phase and restore the database to the state it was in at the time of the crash.
➢ **Undo:** Undo the effects of transactions that failed to commit.

15.   What is a two-phase commit? (U)

Two-phase commit is important whenever a given transaction can interact with several independent "resource managers," each managing its own set of recoverable resources and maintaining its own recovery log."

16. What is the need for save points? (U)

It might be possible for a transaction to establish intermediate save points while it is executing, and subsequently to roll back to a previously established save point, if required, instead of having to roll back all the way to the beginning.

17. Define concurrency (MAY 2012) (R)

Concurrency refers to the fact that DBMSs typically allow many transactions to access the same database at the same time.

18. What are the three problems that any concurrency control mechanism must address? (R)

The three problems are:

> ➢ The lost update problem
> ➢ The uncommitted dependency problem
> ➢ The inconsistent analysis problem

19. What is the last update problem? (U)

Transaction A retrieves some tuple t at time t1; transaction B retrieves that same tuple t at time t2; transaction A updates the tuple at time t3; and transaction B updates the same tuple at time t4; Transaction A's update is lost at time t4, because transaction B overwrites it with-out even looking at it.

20. What is the uncommitted dependency problem? (U)

The uncommitted dependency problem arises if one transaction is allowed to retrieve-or, worse, update-a tuple that has been updated by another transaction but not yet committed by that other transaction.

21. What is deadlock? (NOV 2014) (R)

Deadlock is a situation in which two or more transactions are in a simultaneous wait state, each of them waiting for one of the others to release a lock before it can proceed.

22. What is serializability? (NOV2016) (NOV 2014) (R)

Serializability is the generally accepted "criterion for correctness" for the interleaved execution of a set of transactions; that is, such an execution is considered to be correct if and only if it is serializable.

23.   State the two-phase locking protocol. (MAY 2012)(NOV 2013) (R)

The two-phase locking protocol is as follows

i. Before operating on any object a transaction must acquire a lock on that object.

ii. After releasing a lock, a transaction must never go on to acquire any more locks.

24.   What is an isolation level? (R)

The isolation level that applies to a given transaction might be defined as the degree of interference the transaction in question is prepared to tolerate on the part of concurrent transactions.

25.   What do you mean by phantom problem? (R)

If transactions operate at less than the maximum isolation level is the so-called phantom problem

26.   What is an intent locking protocol? (U)

The intent locking protocol, according to which no transaction is allowed to acquire a lock on a tuple before first acquiring a lock-probably an intent lock on the relvar that contains it.

27.   List the four conditions for Deadlock(U) (NOV 2016)

A deadlock situation can arise if and only if the following four conditions hold simultaneously in a system-

Mutual Exclusion: At least one resource is held in a non-sharable mode that is only one process at a time can use the resource. If another process requests that resource, the requesting process must be delayed until the resource has been released.

Hold and Wait:There must exist a process that is holding at least one resource and is waiting to acquire additional resources that are currently being held by other processes.

No Preemption: Resouces cannot be preempted; that is, a resource can only be released voluntarily by the process holding it, after the process has completed its task.

Circular Wait: There must exist a set $\{p_0, p_1,.....p_n\}$ of waiting processes such that p_0 is waiting for a resource which is held by p_1, p_1 is waiting for a resource which is held by p_2,...,

p_{n-1} is waiting for a resource which is held by p_n and p_n is waiting for a resource which is held by p_0.

28.   Why DBMS needs a concurrency control? NOV 2017

In general, concurrency control is an essential part of Transaction management. It is a mechanism for correctness when two or more database transactions that access the same data or data set are executed concurrently with time overlap. According to Wikipedia.org, if multiple transactions are executed serially or sequentially, data is consistent in a database. However, if concurrent transactions with interleaving operations are executed, some unexpected data and inconsistent result may occur. Data interference is usually caused by a write operation among transactions on the same set of data in DBMS.

29.   What are data fragmentations?State the various fragmentations with example. (R) NOV 2017.

Fragmentation occurs in a dynamic memory allocation system when many of the free blocks are too small to satisfy any request.

External Fragmentation: External Fragmentation happens when a dynamic memory allocation algorithm allocates some memory and a small piece is left over that cannot be effectively used. If too much external fragmentation occurs, the amount of usable memory is drastically reduced. Total memory space exists to satisfy a request, but it is not contiguous.

Internal Fragmentation: Internal fragmentation is the space wasted inside of allocated memory blocks because of restriction on the allowed sizes of allocated blocks. Allocated memory may be slightly larger than requested memory; this size difference is memory internal to a partition, but not being used

30.   Define ordered indices with example (U)NOV 2017

In order to allow fast random access, an index structure may be used.

A file may have several indices on different search keys.

If the file containing the records is sequentially ordered, the index whose search key specifies the sequential order of the file is the primary index, or clustering index. Note: The search key of a primary index is usually the primary key, but it is not necessarily so.

Indices whose search key specifies an order different from the sequential order of the file are called the secondary indices, or non-clustering indices.

31.   What are the states of transaction? (R) (May-2019)

Transaction states are as follows-

II Year/IV Semester/2021-2022 DBMS

1. Active state
2. Partially committed state
3. Committed state
4. Failed state
5. Aborted state
6. Terminated state

32. **What is meant by log based recovery? (R) (May-2019)**

The log is a sequence of records. Log of each transaction is maintained in some stable storage so that if any failure occurs, then it can be recovered from there.

33. **List the ACID properties and its usefulness. (R) (May-2021)**

ACID properties are

(i) Atomicity: Transactions are atomic

(ii) Correctness: Transactions transform a correct state of the database into another correct state, without necessarily preserving correctness at all intermediate points

(iii) Isolation: Transactions are isolated from one another. That is, even though in general there will be many transactions running concurrently, any given transaction's updates all connected from all the rest, until that transaction commits.

(iv) Durability: Once a transaction commits, its updates persist in the database, even if there is a subsequent system crash.

34. **What benefit does strict two-phase locking provide? What are the disadvantages of it? (AN) (May-2021)**

In strict two-phase locking protocol

• Locks are obtained and released in two separate phases just like two-phase locking protocol.

• Additionally it also requires that all the exclusive mode locks should be held by the transaction till it commits.

• Because it produces only cascadeless schedules, recovery is very easy.

• But the set of schedules obtainable is a subset of those obtainable from plain two phase locking, thus concurrency is reduced.

PART – B

II Year/IV Semester/2021-2022 DBMS

1.Why is Recovery needed?Discuss any two Recovery Techniques**.(U)(MAY2009) (MAY 2012)**

2. Write a relevant example discuss two phase Looking. **(U)(MAY 2009)**

3. What is Deadlock? List and discuss the four conditions for Deadlock. **(R)**

(MAY 2009)(MAY 2012) (NOV 2014) (NOV2016)

4.Give an example of a Scenario where two Phase Locking leads to Deadlock.(AN)(May2009)

5. Explain testing for serializability with respect to concurrency control schemes. How

will you determine, whether a schedule is serializble or not. **(U)(MAY 2008) (MAY 2019)**

6. Explain the following protocols for concurrency control :(**R)(MAY 2008)**

   a.   Look based protocols.

   b.   Time stamp based protocols.

7. Describe the three most common concurrent transaction execution problems.**(R)(MAY 2007)**

8. Explain how concurrency control can be used to avoid such problems?(U) **(MAY 2007)**

9. Write a note on the following : **(NOV 2014)**

   c.   Two-phase locking protocol (2PL).(U)(NOV 2016)

   d.   Lock based protocol.(U)

10. Explain the deferred and immediate-modification version of the log based Recovery scheme.

(U)(MAY 2007)

11.Write short notes on shadow paging.(R)(**DEC 2007)**

12.Describe about the testing of serializability. (U)(**DEC 2007) (MAY 2012)**

13.What is concurrency control? How is it implemented in DBMS? Explain. **(R)(DEC 2007)**

14.Explain various recovery techniques during transaction in detail. **(R)(MAY 2007)**

15.Explain the two-phase commit and three-phase commit protocols**.(R)(Dec2015)**

 16**.** State and explain the lock based concurrency control with suitable example.**(R)** (**NOV 2017)**

 17. When does deadlock occurs? Explain two-phase commit protocol with example. **(AN) (NOV 2017)**

 18. Explain deferred and immediate modification versions of the log based recovery schemes.

 (U) (May 2019)

19. State and explain the transaction isolation level.**(U) (MAY 2021)**

20. What are the two approaches of deadlock prevention? Explain in detail with suitable example. **(U) (MAY 2021)**

COURSE OUTCOMES

Apply concurrency control and recovery mechanisms for practical problems and design the Query Processor and Transaction Processor

UNIT IV

IMPLEMENTATION TECHNIQUES

SYLLABUS: RAID – File Organization – Organization of Records in Files – Indexing and Hashing –Ordered Indices – B+ tree Index Files – B tree Index Files – Static Hashing – Dynamic Hashing – Query Processing Overview – Algorithms for SELECT and JOIN operations – Query optimization using Heuristics and Cost Estimation.

PART – A

1. **What is cache memory?(U)**

   The cache is the fastest and most costly form of storage. Cache memory is small; its use is managed by the computer system hardware.

2. **What is main memory?(U)**

   The storage medium used for data that are available to be operated on is main memory. The general-purpose machine instructions operate on main memory. Although main memory may contain many megabytes of data, or even hundreds of gigabytes of data in large server systems, it is generally too small for storing the entire database. The contents of main memory are usually lost if a power failure or system crash occurs.

3. **What is flash memory? (MAY 2010)(U)**

   EEPROM (Electrically Erasable Programmable Read Only Memory)

a. Data in flash memory survive from power failure

b. Reading data from flash memory takes about 10 nano-secs (roughly as fast as from main memory), and writing data into flash memory is more complicated: write-once takes about 4-10 microsecs.

c. To overwrite what has been written, one has to first erase the entire bank of the memory. It may support only a limited number of erase cycles (10^4 to 10^6).

d. It has found its popularity as a replacement for disks for storing small volumes of data (5-10 megabytes).

4. What is magnetic disk storage? (U)

Primary medium for long-term storage.

➢ Typically the entire database is stored on disk.

➢ Data must be moved from disk to main memory in order for the data to be operated on.

➢ After operations are performed, data must be copied back to disk if any changes were made.

➢ Disk storage is called direct access storage as it is possible to read data on the disk in any order (unlike sequential access).

➢ Disk storage usually survives power failures and system crashes

5. What is optical storage?(U)

The most popular form of optical storage are CD-ROM (compact – disk read – only memory), WORM (write-once read-many)disk (for archival storage of data), and Juke box(containing a few drives and numerous disks loaded on demand).

6. What is tape storage?)(U)

Tape storage used for primarily for backup and archival data.

➢ Cheaper, but much slower access, since tape must be read sequentially from the beginning.

➢ Used as protection from disk failures.

7. What is SAN architecture?(R)

In the storage area network (SAN) architecture, large numbers of disks are connected by a high-speed network to a number of server computers

8.   What is NAS?(R)

II Year/IV Semester/2021-2022 DBMS

Network attached storage (NAS) is an alternative to SAN. NAS is much like SAN, except that instead of the networked storage appearing to be a large disk, it provides a file system interface using networked file system protocols such as NFS or CIFS.

9.   Draw the storage device hierarchy. (MAY 2010)(R)
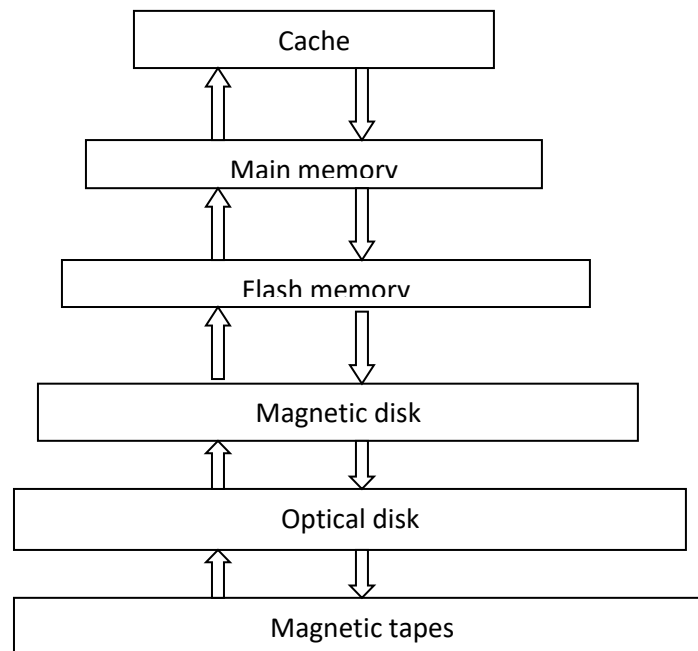
10. Define access time(R)

Access time is the time from when a read or write request is issued to when data transfer begins.

11. Define seek time.(R)

To access data on a given sector of a disk, the arm first must move so that it is positioned over the correct track, and then must wait for the sector to appear under it as the disk rotates. The time for repositioning the arm is called the seek time.

12. Define average seek time.(R)

The average seek time is the average of the seek times, measured over a sequence of random requests.

13. Define rotational latency time.(U)

Once the head has reached the desired track, the time spent waiting for the sector to be accessed to appear under the head is called the rotational latency time

14. Define average latency time.(U)

Average latency time of the disk is one-half the time for a full rotation of the disk.

15. Define data-transfer rate.(U)

The data-transfer rate is the rate at which data can be retrieved from or stored to the disk.

16. What is RAID? (NOV 2014)(R)

A variety of disk organization techniques, collectively called redundant arrays of independent disks (RAID), have been proposed to achieve improved performance and reliability.

17. What are the factors to be taken into account in choosing a RAID level?(R)

The factors to be taken into account in choosing a RAID level are

- Monetary cost of extra disk-storage requirements
- Performance requirements in terms of number of I/O operations
- Performance when a disk has failed.
- Performance during rebuild(that is, while the date in a failed disk are being rebuilt on a new disk)

18. What is heap file organization?(R)

Any record can be placed anywhere in the file where there is space for the record. There is no ordering of records. Typically, there is a single file for each relation.

19. What is sequential file organization?(R)

Records are stored in sequential order, according to the value of a "search key" of each record.

20. What is hashing file organization?(R)

II Year/IV Semester/2021-2022 DBMS

A hash function is computed on some attribute of each record. The result of the hash function specifies in which block of the file the record should be placed,

21. What is a multitable clustering file organization?(U)

A multitable clustering file organization is a file organization that stores related records of two or more relations in each block. Such a file organization allows us to read records that would satisfy the join condition by using one block read.

22. What are the two basic kinds of indices?(U)

The Two basic kinds of indices are

- ➢ Ordered indices
- ➢ Hash indices

23. What are access types?(U)

The types of access that are supported efficiently. Access types can include finding records with a specified attribute vale and finding records whose attribute values fall in a specified range.

24. What is access time?(U)

The time it takes to find a particular data item, or set of items, using the technique in question.

25. What is insertion time?(U)

The time it takes to insert a new data item. This value includes the time it takes to find the correct place to insert the new data item, as well as the time it takes to update the index structure.

26. What is deletion time?(U)

The time it takes to delete a data item. This value includes the time it takes to find the item to be deleted, as well as the time it takes to update the index structure.

27. What is space overhead?(U)

The additional space occupied by an index structure. Provided that the amount of additional space is moderate, it is usually worth-while to sacrifice the space to achieve improved performance.

II Year/IV Semester/2021-2022 DBMS

28. What is a search key?(R)

An attribute or set of attributes used to look up records in a file is called a search key.

29. What is an clustering index (or) primary index? (NOV 2014)(U)

Clustering index is an index whose search key also defines the sequential order of the file.

30. What is an nonclustering index (or) secondary index?(U)

Indices whose search key specifies an order different from the sequential order of the file are called nonclustering indices.

31. What is an index-sequential file?(U)

All files are ordered sequentially on some search key. Such files, with a clustering index on the search key, are called index-sequential files.

32. What is a dense index? (U) (MAY 2012) (MAY 2019)

An index record appears for every search-key value in the file. In a dense clustering index, the index record contains the search-key value and a pointer to the first data record with that search-key value. The rest of the records with the same search-key value would be stored sequentially after the first record, since, because the index is a clustering one, records are stored on the same search key. Dense index implementations may store a list of pointers to all records with the same search-key value; doing so is not essential for clustering indices.

33. What is a sparse index? (MAY 2012)(U)

An index record appears for only some of the search-key values. as is true in dense indices, each index record contains a search-key value and a pointer to the first data record with that search-key value. To locate a record, we find the index entry with the largest search-key value that is less than or equal to the search-key value for which we are looking. We start at the record pointed to by that index entry, and follow the pointers in the file until we find the desired record.

34. What is a multilevel indices?(U)

Indices with two or more levels are called multilevel indices.

35. What is a balanced tree?(U)

Balanced tree in which every path from the root of the tree to the leaf of the tree is of the same length.

36. Write the querying a B+ tree.(AP)

Procedure find (value V)

    Set C = root node

    While C is not a leaf node begin

      Let K_i= smallest search-key value, if any, greater than V

      If there is no such value then begin

        Let m = the number of pointers in the node

        Set C= node pointed to by P_m

      End

        Else set C= the node pointed to by P_i

    End

    If there is a key value K_i =V

      Then pointer P_i directs us to the desired record or bucket

      Else no record with key value k exists.

37. Define hashing.(R)

Hashing allow us to avoid accessing an index structure. It provides a way of constructing indices.

38. Define a bucket.(R)

The term bucket to denote a unit of storage that can store one or more records. A bucket is typically a disk block, but could be chosen to be a smaller or larger than a disk block.

39. Define hash function.(U)

A hash function h is a function from K to B. Let h denote a hash function

40. What are the two different purposes of hashing?(R)

Hashing can be used for two different purposes.

- Hash file organization
- Hash index organization

41. What is a hash index?(R)

A hash index organizes the search keys, with their associated pointers, into a hash file structure.

42. What is dynamic hashing? (NOV 2014)(R)

Dynamic hashing techniques allow the hash function to be modified dynamically to accommodate the growth or shrinkage of the database.

43. What are the steps involved in query processing?(R)

The steps involved in processing a query are

- Parsing and translation
- Optimization
- Evaluation

44. Write the nested-loop join algorithm.(U)

For each tuple t_r in r do begin

      For each tuple t_s in s do begin

        Test pair (t_r, t_s) to see if they satisfy the join condition

         If they do, add $tr.t_s$ to the result

         End

        End

45. Write the block nested-loop join and its algorithm.(U)

Block nested-loop join, which is a variant of the nested – loop join where every block of the inner relation is paired with every block of the outer relation.

      For each block B_r of r do begin

     For each block B_s of s do begin

      For each tuple t_r in B_r do begin

       For each tuple t_s in B_s do begin

        Test pair (t_r, t_s) to see if they satisfy the join condition

        If they do, add $t_r.t_s$ to the result.

        End

         End

        End

End

46. What is an indexed nested-loop join?(U)

Indexed nested loop join can be used with existing indices, as well as with temporary indices created for the sole purpose of evaluating the join.

47. What is a merge join?(U)

The merge-join algorithm can be used to compute natural joins and equi-joins.

48. What is a hash-join?(U)

The hash-join algorithm can be used to implement natural joins and equi-joins.

49.   What are the contents of DBMS catalog?(R)

The DBMS catalog stores the following statistical information about database relations:

- n_r,the number of tuples in the relation r
- b_r, the number of blocks containing tuples of relation r
- l_r, the size of a tuple of relation r in bytes
- f_r,the blocking factor of relation r-that is,the number of tuples of relation r that fit into one block
- $V(A,r)$,the number of distinct values that appear in the relation r for attribute A. This value is the same as the size of (r).If A is a key for relation r,$V(A,r)$is n_r

50. What is a slotted page sheet.draw the diagram(NOV 2013)(R)

Record id = <page id, slot #>

Can move records on page without changing rid; so, attractive for fixed-length records too.

Page is full when data space and slot array meet.
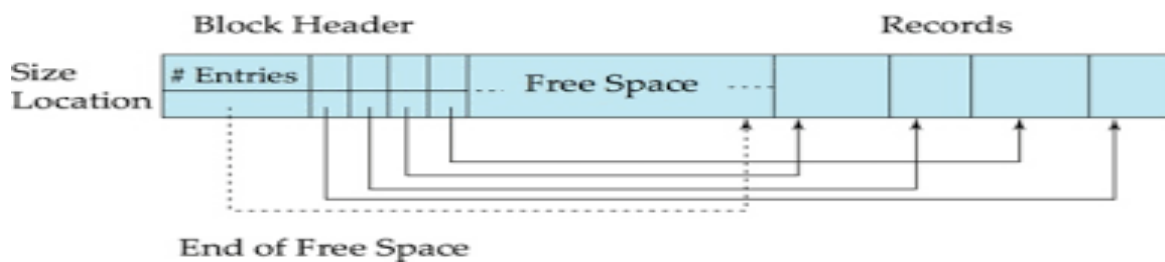
51. What is a histogram?(U)

In histogram the values for the attribute are divided into a number of ranges, and with each range the histogram associates the number of tuples whose attribute value lies in that range.

52. What is an index record?(U)

II Year/IV Semester/2021-2022 DBMS

An index record, or index entry, consists of a search – key vale and pointers to one or more records with that value as their search-key value.

53. Define mean time to failure(MTTF)(R)

Mean time to failure is a measure of the reliability of the disk. The mean time to failure of a disk is the amount of time that, on average, we can expect the system to run continuously without any failure.

54. What is mirroring?(R)

The simplest approach to introducing redundancy is to duplicate every disk. This technique is called mirroring.

55.How dynamic hashing differ from static hashing?(DEC2015)(U)

Static Hashing

- A bucket is a unit of storage containing one or more records (a bucket is typically a disk block).
- The file blocks are divided into M equal-sized buckets, numbered bucket0, bucket1... bucketM-1.Typically, a bucket corresponds to one (or a fixed number of) disk block.
- In a hash file organization we obtain the bucket of a record directly from its search-key value using a hash

Dynamic Hashing

- Good for database that grows and shrinks in size
- Allows the hash function to be modified dynamically
- **Extendable hashing** – one form of dynamic hashing
- Hash function generates values over a large range —typically *b*-bit integers, with $b = 32$.
- At any time use only a prefix of the hash function to index into a table of bucket addresses.
- Let the length of the prefix be *i* bits, $0 \_ i \_ 32$.

55. Disadvantages of B TREE over B+ TREE(R)(NOV 2016)

In a B tree search keys and data stored in internal or leaf nodes. But in B+-tree data store only leaf nodes.

Searching any data in a B+ tree is very easy because all data are found in leaf nodes. In a B tree, data cannot be found in leaf nodes.

In a B tree, data may be found in leaf nodes or internal nodes. Deletion of internal nodes is very complicated. In a B+ tree, data is only found in leaf nodes. Deletion of leaf nodes is easy.

Insertion in B tree is more complicated than B+ tree.

B+ trees store redundant search key but B tree has no redundant value.

In a B+ tree, leaf nodes data are ordered as a sequential linked list but in B tree the leaf node cannot be stored using a linked list. Many database systems' implementations prefer the structural simplicity of a B+ tree.

56. Mention all the operations of file.(R) (MAY 2019)

File Operations

OPEN, LOCATE, READ, WRITE, CLOSE

57. State the storage device hierarchy. (R) (MAY 2021)
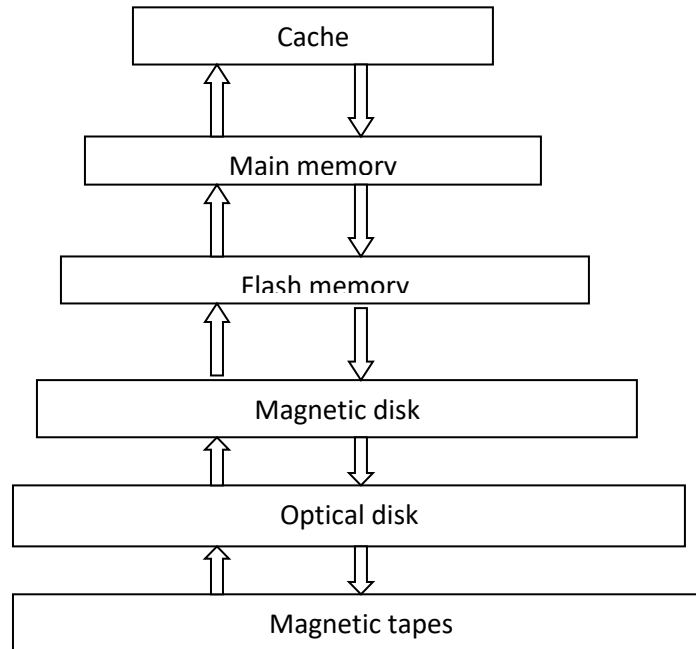
58. What are the factors needed to evaluate the technique of ordered indexing and hashing? (R) (MAY 2021)

Indexing and Hashing techniques are evaluated on the basis of

Access types: The types of access that are supported efficiently. Access types can include finding records with a specified attribute value and finding records whose attribute values fall in a specified range.

Access time: The time it takes to find a particular data item, or set of items, using the technique in question.

II Year/IV Semester/2021-2022 DBMS

Insertion time: The time it takes to insert a new data item.

Deletion time: The time it takes to delete a data item.

Space overhead: The additional space occupied by an index structure.

PART B

1. Explain the security features provided in commercial query languages.(U) **(MAY 2007)**

2. What are the steps involved in query processing? How would you estimate the cost ofthe query? **(MAY 2007)(U)**

3. Explain the different properties of indexes in detail.(U)

4. Explain various hashing techniques.(R)

5. Explain the four important properties of transaction that a DBMS must ensure to maintain database.(U)**(MAY 2007)**

6. What is RAID? List the different level in RAID technology and explain its Features.(R)**(MAY 2007), (MAY 2010) (NOV 2014)(NOV 2016) (MAY 2019)**

7. What is concurrency control? How is it implemented in DBMS? Explain**.(U)(MAY 2007)**

8. Explain various recovery techniques during transaction in detail.(U) (**MAY 2007**)

9. Explain how the RAID systems improve performance and reliability**. (U)(DEC 2007**)

10. Describe the structure of B^+ tree and list the characteristics of a B^+ tree**.(U)**

                                      (DEC 2008), **(MAY 2012) (MAY 2010) (NOV 2014)**

11. Explain the steps involved in processing a query**.(R)(DEC 2008)**

12. Describe static hashing and dynamic hashing**.(U)(DEC 2008). (NOV 2014)**

13. Describe in detail about how records are represented in a file and how to organize them in a file**. (AP)(MAY 2012).**

14. Explain about spatial and mobile database (U)**(NOV 2014)(NOV 2016)**

15. With suitable digrams,discuss about the Raid Levels(Level0,Level1,Level 0+1,level3,Level and Level5)**.(U)(Dec2015)**

16. Explain the Architectural components of a data ware house and write bout data marts.(R)(**Dec2015**)

17. Explain the architecture distributed database system **(U)(Nov 2016)**

18. Describe the benefits and drawbacks of a source driven architecture for gathering of data ata

datawarehouse , as compared to a destination driven architecture**(U)(NOV 2016)**

19. i) What are the various feature of distributed database versus centralized database system ?

ii) Explain the B+ tree indexes on multiple keys with a suitable example. **(AP)(NOV 2017)**

20. Explain the distinction between static and dynamic hashing. Discuss the relative merits of

each technique in database applications**. (AN) (NOV 2017)**

21. Describe the structure of B+ Tree and give the algorithm for searching in B+ tree with example. **(U) (MAY 2019)**

22. Describe the procedure for index update for single level indices with example. **(U) (MAY 2021)**

23. Explain dynamic hashing with example. **(U) (MAY 2021)**

COURSE OUTCOMES

Able to select best data storage medium for databases.

UNIT V

ADVANCED TOPICS

SYLLABUS: Distributed   Databases: Architecture,   Data   Storage,   Transaction Processing – Object-based Databases: Object Database Concepts, Object-Relational features, ODMG Object Model, ODL, OQL - XML Databases: XML Hierarchical Model, DTD, XML Schema, XQuery – Information Retrieval: IR Concepts, Retrieval Models, Queries in IR systems.

PART-A

1. **What are the different aspects of security problem?(R)**

There are many aspects to the security problem. Here are some of them:

       i. Legal, social, and ethical aspects

      ii. Physical controls

II Year/IV Semester/2021-2022 DBMS

       iii.   Policy questions

       iv.   Operational problems

       v.   Hardware controls

       vi.   Operational system support

       And finally

➢ Issues that are the specific concern of the database system itself.

2. What is discretionary access control?(U)

     In the case of discretionary control, a given user will typically have different access rights on different objects; further, there are few inherent limitations regarding which users can have which rights on which objects. Discretionary schemes are thus flexible.

3. What is mandatory access control?(U)

     In the case of mandatory control, by contrast, each data object is labeled with a certain classification level, and each user is given a certain clearance level.

4. What are the contents of audit trail record?(R)

     A typical audit trail record might contain the following information:

➢ Request(source text)

➢ Terminal from which the operation was invoked

➢ User who invoked the operation

➢ Date and time of the operation

➢ Relvar(s), tuples(s),attribute(s) affected

➢ Before images(old values)

➢ After images(new values)

5. What is entity integrity?(U)

     No component of the primary key of any base relvar is allowed to accept nulls are called entity integrity.

6. What are views for?(U)

     There are many reasons why view support is desirable. Here are some of them:

➢ Views provide a shorthand or "macro" capability.

II Year/IV Semester/2021-2022 DBMS

> ➢ Views allow the same data to be seen by different users in different ways at the same time.
>
> ➢ Views provide automatic security for hidden data.
>
> ➢ Views can provide logical data independence.

7. Define the golden rule.(R)

Golden Rule is

No update operation must ever assign to any relvar a value that causes its relvar predicate to evaluate to FALSE.

Or (a little loosely):

No relvar must ever be allowed to violate its own predicate.

8. Define a distributed database system.(U)

A distributed database system consists of a collection of sites, connected together via some kind of communications network, in which:

    a. Each site is a full database system site in its own right but

    b. The sites have agreed to work together so that a user at any site can access data anywhere in the network exactly as if the data were all stored at the user's own site.

9. Define a distributed database management system.(R)(NOV 2016)

A new software component at each site logically an extension of the local DBMS provides the necessary partnership functionality, and it is the combination of these new components together with the existing DBMSs that constitutes what is usually called the distributed database management system

10. What are the advantages of distributed databases?(U)

It enables the structure of the database to mirror the structure of the enterprise-local data can be kept locally, where it most logically belongs-while at the same time remote data can be accessed when necessary.

11. What is the fundamental principle of distributed database?(U)

The fundamental principle of distributed database is to the user, a distributed system should look exactly like a non-distributed system.

12. What are the objectives of distributed databases?(R)

The objectives of distributed databases are

    1) Local autonomy

    2) No reliance on a central site

    3) Continuous operation

    4) Location independence

    5) Fragmentation independence

    6) Replication independence

    7) Distributed query processing

    8) Distributed transaction management

    9) Hardware independence

    10) Operating system independence

    11) Network independence

    12) DBMS independence

13. What are the problems of distributed databases?(R)

The problems are

- ➢ Query processing
- ➢ Catalog management
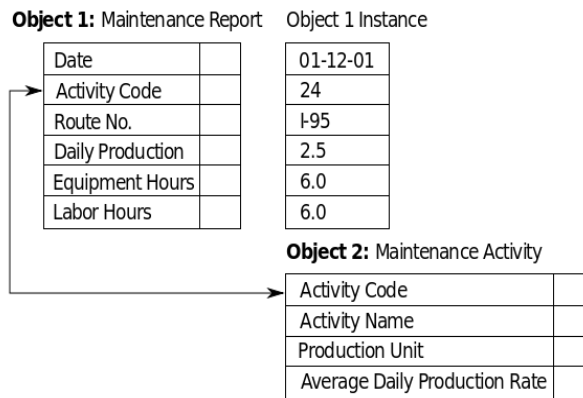- ➢ Update propagation
- ➢ Recovery
- ➢ Concurrency

14. Define client/server systems.(R)

A client/server system is a distributed system in which

a) Some sites are client sites and some are server sites

b) All data resides at the server sites

c) All applications execute at the client sites, and

d) "the seams show

15. What is OODBMS ?(U)

Object-oriented database management systems (OODBMSs) combine database capabilities with object-oriented programming language capabilities. OODBMSs allow object-oriented programmers to develop the product, store them as objects, and replicate or modify existing objects to make new objects within the OODBMS. Because the database is integrated with the programming language, the programmer can maintain consistency within one environment, in that both the OODBMS and the programming language will use the same model of representation.

Object-Oriented Model

Object 1: Maintenance Report   Object 1 Instance

Date	
Activity Code	
Route No.	
Daily Production	
Equipment Hours	
Labor Hours	

01-12-01
24
I-95
2.5
6.0
6.0

Object 2: Maintenance Activity

Activity Code	
Activity Name	
Production Unit	
Average Daily Production Rate	

16. What is meant by Data warehousing? (NOV 2014)(U)

A data warehouse is a relational database that is designed for query and analysis rather than for transaction processing. It usually contains historical data derived from transaction data, but it can include data from other sources. It separates analysis workload from transaction workload and enables an organization to consolidate data from several sources.

17. What is meant by data mining? (NOV 2014)(R)

The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

18. What is meant by information retrieval?(R)

An information retrieval process begins when a user enters a query into the system. Queries are formal statements of information needs, for example search strings in web search engines. In information retrieval a query does not uniquely identify a single object in the collection. Instead, several objects may match the query, perhaps with different degrees of relevancy.

19. What is meant by Relevancy ranking? (NOV 2014)(U)

Relevancy ranking is the process of sorting the document results so that those documents which are most likely to be relevant to your query are shown at the top.

20. Define Crawling. (NOV 2014)(R)

Web crawling is the process of search engines combing through web pages in order to properly index them. These "web crawlers" systematically crawl pages and look at the keywords contained on the page, the kind of content, all the links on the page, and then returns that information to the search engine's server for indexing. Then they follow all the hyperlinks on the

II Year/IV Semester/2021-2022 DBMS

website to get to other websites. When a search engine user enters a query, the search engine will go to its index and return the most relevant search results based on the keywords in the search term. Web crawling is an automated process and provides quick, up to date data.

21. What is meant by XML Database?(R)

An XML database is a data persistence software system that allows data to be specified, and sometimes stored, in XML format. These data can then be queried, transformed, exported and returned to a calling system.

22. Define Data classification.(U)

Data classification is the process of organizing data into categories for its most effective and efficient use.

23. What is meant by Access Control? (NOV 2014)(R)

In the fields of physical security and information security, access control is the selective restriction of access to a place or other resource. The act of accessing may mean consuming, entering, or using. Permission to access a resource is called authorization. Locks and login credentials are two analogous mechanisms of access control.

24. What is a threat in dbms?(U)

When the data is been accessed by many people, the chances of data theft increases. In the past, database attacks were prevalent, but were less in number as hackers hacked the network more to show it was possible to hack and not to sell proprietary information. Another reason for database attacks is to gain money selling sensitive information, which includes credit card numbers, Social Security Numbers, etc.

25. What is Cryptography?(U)

Cryptography is a method of storing and transmitting data in a particular form so that only those for whom it is intended can read and process it. The term is most often associated with scrambling plaintext (ordinary text, sometimes referred to as cleartext) into ciphertext (a process called encryption), then back again (known as decryption).

26. Define Statistical database.(R)

A statistical database is a database used for statistical analysis purposes. It is an OLAP (online analytical processing), instead of OLTP (online transaction processing) system. Modern decision, and classical statistical databases are often closer to the relational model than the multidimensional model commonly used in OLAP systems.

27. Define Clustering in DBMS. (NOV 2014)(R)

II Year/IV Semester/2021-2022 DBMS

Clustering, in the context of databases, refers to the ability of several servers or instances to connect to a single database. An instance is the collection of memory and processes that interacts with a database, which is the set of physical files that actually store data.

28. Write about the four types (Star, Snowflake, Galaxy and Fast constellation) of Data warehouse schemas**.(DEC2015)(R)**

1.STAR SCHEMA:Centralized Fact table connect the one or more denormalized data

2.SNOW FLAKE SCHEMA:Centralized Fact table connect the one or more normalized data

3.STAR FLAKE SCHEMA:One or more centralized fact table connect the single denormolized data

4. Fact Constallation Schema : For each star schema it is possible to construct fact constellation schema(for example by splitting the original star schema into more star schemes each of them describes facts on another level of dimension hierarchies). The fact constellation architecture contains multiple fact tables that share many dimension tables.

29.What is Association rule mining?(R)

Association rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational database or other information repository. An example of an association rule would be "If a customer buys a dozen eggs, he is 80% likely to also purchase milk."

Applications:

- Cross-Marketing
- Basket Data Analysis
- Catalog design

30. Define XQuery (U)

XQuery is the query language used to query the records in the XML document. It is similar to the SQL query in RDBMS. It is the functional language independent of the format of the XML data in the document. This can be used to query XML document, XML database or relational database with XML data.

31. List out the characteristics of XQuery. (R)

II Year/IV Semester/2021-2022 DBMS

- It is a functional language used to retrieve the XML data from any document or database.
- This is similar to SQL in RDBMS, and queries / traverses XML data.
- It uses XPaths to traverse through XML data to fetch the records.
- XQuery is supported by almost all database system.

Though the expressions that we have seen above to point to different nodes, simply writing the expression will not give the required results. Those expressions will simply point to the nodes specified by the expressions. It should have proper querying to pull the values at that node. That is done by XQuery.

32. Define ODMG Object Model (U)

The ODMG Object Model is intended to allow portability of applicationsamong object database products. It provides a common model for these productsby defining extensions to the OMG object model that support object databaserequirements. In particular, the ODMG model extends the OMG core to providefor persistent objects, object properties, more specific object types,queries and transactions.

33. Defining a simple schema with ODL(U)

The first step is to define the database schema. A standard example in databases is the well known Person class (or table in relational system) which contains a few attributes such as a firstname, a lastname, an age, an address, a spouse and a set of children.

We will show the inheritance feature through the simple class Employee which inherits from the Person class and will contains a simple attribute: salary.

Here is simple ODL schema for the classes Address, Person and Employee:

```
//
// person.odl
//
class Address {
  int num;
  string street;
  string town;
  string country;
};
```

```
class Person {
 string firstname;
 string lastname;
 int age;
 Address addr;
 Person * spouse inverse Person::spouse;
 set<Person *> children;
};
class Employee extends Person {
 long salary;
};
```

34. What is OQL (U)

OQL is the way to access data in an O2 database. OQL is a powerful and easy-to-use SQL-like query language with special features dealing with complex objects, values and methods.

Using OQL

- Setup your environment
- SELECT, FROM, WHERE
- Dot notation and path expressions
- Subqueries in FROM clause
- Subqueries in WHERE clause
- Set operations and Aggregation
- GROUP BY
- Embedded OQL

35. differences between OQL and SQL (A)

Network Manager uses OQL to transfer data between and communicate with its internal databases. OQL is an object-based version of Structured Query Language (SQL) that was designed specifically around the operational needs of the Network Manager architecture.

The following items identify the main differences between OQL and SQL:

- OQL has the ability to support object referencing within database tables. Thus, it is possible to have objects nested within objects.

II Year/IV Semester/2021-2022 DBMS

- Not all SQL keywords are supported within OQL. Thus, irrelevant keywords have been removed for the OQL syntax.

36. List out the types of Queries in IR Systems (R)

Keyword Queries. Keyword-based queries are the simplest and most commonly used forms of IR queries: the user just enters keyword combinations to retrieve documents.

- Boolean Queries. ...
- Phrase Queries. ...
- Proximity Queries. ...
- Wildcard Queries. ...
- Natural Language Queries.

37. What are the Difference between an IT Security Vulnerability, Threat and Risk?
 (AN) (NOV2017)

Threat

A threat refers to a new or newly discovered incident with the potential to do harm to a system, network, firewalls or your overall organization.

There are three main types of threats:

Natural threats (e.g., floods or a tornado),

Unintentional threats (such as an employee mistakenly accessing the wrong information) Intentional threats. There are many examples of intentional threats including spyware, malware, adware companies or the actions of a disgruntled employee. In addition, worms and viruses are also categorized as threats, because they could potentially cause harm to your organization through exposure to an automated attack, as opposed to one created by humans.

Vulnerability

A vulnerability refers to a known weakness of an asset or devices that can be exploited by one or more attackers. In other words, it is a known issue that allows an attack to be successful. For example, when a team member resigns and you forget to disable their access to external accounts, change logins or remove their names from company credit cards, this leaves your business open to both intentional and unintentional threats.However, most vulnerabilities are exploited by automated attackers and not a human typing on the other side of the network.

Risk

Risk refers to the potential for loss or damage when a threat exploits a vulnerability. Examples of risk include financial losses as a result of business disruption, loss of privacy, reputational damage and can even include loss of life as well.

Risk can also be defined as follows:

Risk = Threat X Vulnerability

Improve Security in overall network and system based.

38. State the XML Schema Functions (R) (NOV2017)

The Virtuoso interface to XML Schema is represented primarily by two functions:

xml_validate_schema()

xml_load_schema_decl()

The signature of the function xml_validate_schema() is the same as the function xml_validate_dtd() . It parses and validates an XML document. The root element of the document must contain the"schemaLocation" attribute with the value of the document's URI.

As described above, the XML Schema Processor implemented within Virtuoso relies on the XML Schema DTD, which is composed of two files: "XMLSchema.dtd" and "datatypes.dtd." These files must be placed in the system directory (see xml_add_system_path()).

The following XML Schema items are not fully implemented:

•   facets support is primitive;

•   you may only derive by restriction from the "anyType" type;

•   enumerations are not supported;

•   the "all" particle is not supported;

•   elements may not be defined within an element model group declaration;

•   unions are not supported;

•   "appinfo," "documentation," "list," and "notation" tags are ignored.

Virtuoso does not cache XML Schema documents; they are completely reprocessed every time the document is loaded.

39. Mention two features of multimedia database. (R) (MAY 2019)

Corresponding storage media: Multimedia data must be stored and managed according to the specific characteristics of the available storage media.

Comprehensive search methods: During a search in the database, an entry, given in the form of text or a graphical image, is found using different search queries and the corresponding search methods.

II Year/IV Semester/2021-2022 DBMS

Format independent interface: database queries should be independent of media format. MDBMS should provide information in formats requested by the application.

40. Compare Sequential Access Device versus Random Access Devices in DBMS. (R) (MAY 2019)

Random access refers to the ability to access data at random. The opposite of random access is sequential access. To go from point A to point Z in a sequential-access system, you must pass through all intervening points. In a random-access system, you can jump directly to point Z. Disks are random access media, whereas tapes are sequential access media.

41.What is the difference between a false positive and false drop? (R) (MAY 2021)

All tests can potentially result in false positives and false negatives. A false positive (+) describes that the results states you have the condition that were tested for, but you don not really have it. A false negative (-) means that the results states that you do not have a condition, but you actually do.

PART B

1. Explain in detail Data Classification in DBMS.**(R)**

2. Discuss in detail Access Control mechanism. (NOV 2014) **(U)**

3. Explain about Distributed Databases. (NOV 2014) **(U) (MAY 2019)**

4. Discuss about Data warehousing and Data Mining. **(U)**

5. Explain in detail Information retrieval and Relevance Ranking. **(U)**

6. Describe about OODBMS and XML Database. **(R)**

7. Explain in detail Threats and risks in Database Management System. **(R)**

8. Neatly write the K-means algorithm and show the intermediate results in clustering the below given points into two clusters using K-means algorithm. (**Dec2015) (U)**

   P1☺0,0),P2☺1,10),P3:(2,20),P4:(1,15),P5:(1000,2000),P6:(1500,1500),P7:(1000,1250).

9. Discuss About The Access Control Mechanisms And Cryptography Methods To Secure

   the Databases. (**Dec2015) (U)**

II Year/IV Semester/2021-2022 DBMS

10. What is Association rule mining?(**R**)(**NOV 2016**)

11.   Suppose that you have been hired as a consultant to choose a database system for your

client's application .For each of the following applications, state what type of database

system (relational,persistent programming language based OODB,object relational;do not

specify a commercial product)you whould recommend.Justify your recommendation

i) A computer aided design system for manufacturer of airplanes

ii)A system to track contributions made to candiates for public office.

iii)an information system to support the making of movies. **(AN)**(**Nov 2016**)

12. Distinguish between classification and clustering with example. ii) State the necessary for crawling and indexing the web. Explain the procedure for it. **NOV2017**

13. Describe the various component of data warehouse and explain the different data model used to store data with example. **(U)** (**NOV2017)**

14. Discuss in detail about the Deductive DB and Spatial DB**. (R) (MAY 2019)**

15. What are the reasons for building distributed database? Discuss the relative advantages of centralized and distributed databases. Explain the difference between fragmentation, replication and location transparency. **(AN) (MAY 2021)**

16.State and explain the persistent programming languages. **(U) (MAY 2021)**

COURSE OUTCOMES

Apply security concepts to databases

COURSE OUTCOMES

Panimalar Institute of Technology   Department of Information Technology

Course Name : CS8492- DATABASE MANAGEMENT SYSTEMS

Year/Sem   : II /IV

Year of Study : 2021 – 2022 (EVEN)

On Completion of this course, student will gain

CO. NO	DESCRIPTION
C212.1	Design Databases for applications and use the Relational model, ER diagrams.
C212.2	Construct DDL, DML, TCL and DCL commands for different databases with various constraints.
C212.3	Apply concurrency control and recovery mechanisms for practical problems and design the Query Processor and Transaction Processor.
C212.4	Able to select best data storage medium for databases.
C212.5	Apply security concepts to databases

CO-PO-PSO MATRIX

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
C212.1	3	3	3	3	-	-	-	-	-	-	-	2	3	-	-
C212.2	3	3	3	3	-	-	-	-	-	-	-	3	3	-	-
C212.3	3	3	3	3	-	-	-	-	-	-	-	3	3	-	-
C212.4	3	3	3	3	-	-	-	-	-	-	-	2	3	-	1
C212.5	3	3	3	3	-	-	-	-	-	-	-	3	3	-	1
C212	3	3	3	3	-	-	-	-	-	-	-	2.6	3	-	1

II Year/IV Semester/2021-2022 DBMS