Two Marks Questions

1. Define DBMS

    DBMS is Data Base Management System, which consists of a collection of interrelated data and a set of programs to access those data. The collection of data is database. These are designed to manage large bodies of information.

2. Give the disadvantages of file processing system.
    a. Data redundancy and inconsistency
    b. Difficult in accessing data
    c. Data isolation
    d. Integrity problems
    e. Concurrent access anomalies
    f. Security problems

3. What are the three levels of abstraction?
    a. Physical level
    b. Logical level
    c. View level

4. What is physical, logical and view level data abstraction.
    a. The lowest level of abstraction that describes how the data is stored is physical level data abstraction.
    b. The next higher level of abstraction that describes what data are stored in the database and the relationship among those data is logical level data abstraction.
    c. The highest level of abstraction describes only part of the entire database is view level.

5. What do you mean by instances and schemas.

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

    The overall design of the database is called the database schema.

6. What are the two levels of independence?
    a. Physical data independence
    b. Logical data independence

7. What is physical and logical data independence?

    Physical data independence is the ability to modify the physical schema without causing application programs to be rewritten. Modification at the physical level occasionally necessary to improve the performance.

    Logical data independence is the ability to modify the logical schema without causing application programs to be rewritten.

8. Define data model.

    A collection of conceptual tools for describing data, data relationships, data semantics and consistency constraints for underlying the structure of a database is data model.

9. What is DDL?

    Data Definition Language is a special language to define the data base schema. The result of compilation of DDL statements is a set of tables that is stored in a special file called data dictionary or data directory.

10. What are the uses of DML?

    Data Manipulation Language is used to retrieve information stored in the database, insert new record in the database, delete the information from the database and modify the available information in the database.

11. Give the different types of DML.

    Procedural DML requires user to specify what data are needed and how to get those data.

    Non-procedural DML request user to specify what data are needed without specifying how to get the data.

12. Define query.

It is a statement requesting the retrieval of information. The portion of a DML that involves information retrieval is called a query language.

13. What is a transaction?

A transaction is a collection of operations that performs a single logical function in a database application. Each transaction is a unit of both atomicity and consistency.

14. What are the basic notations available in E-R model?
   a. Entity sets.
   b. Relationship sets
   c. Attributes.

15. What do you mean by simple and composite attribute.

The attributes that can be divided into further is simple (eg. Customer city). The attributes that have subparts is composite attribute(eg. Customer name with first name, middle name and last name).

16. Write the two categories of the query language.
   a. Procedural
   b. Non Procedural

17. Define Procedural and non-procedural language.

In a procedural language, the user instructs the system to perform a sequence of operations on a database to complete the desired result.

In a non-procedural language, the user describes the information desired without giving a specific procedure for obtaining that information.

18. List out the six fundamental operators and 4 additional operators in relational algebra.

Six Fundamental operators:

Selection (σ)
Projection (π)
Union ($\cup$)
Set Difference (-)
Cartesian Product (X)
Rename (ρ)

Four Additional operators:

Set Intersection ($\cap$)
Natural Join ($\bowtie$)
Division ($\div$)
Assignment ($\leftarrow$)

19. Which operators are called as unary operators and explain why they are called so.

Unary operators:

Selection (σ)
Projection (π)
Rename (ρ)
These operators are called as binary operators because they operate on only one relation.

20. Which operators are called as binary operators and explain why they are called so.

Binary operators:

Union ($\cup$)
Set Difference (-)
Cartesian Product (X)
These operators are called as binary operators because they operate on pairs of relations.

21. Write a relational algebra expression to find those customers who live in "Harrison"

$$\pi_{customer\_name} (\sigma_{customer\_city="Harrison"} (customer))$$

22. Explain the use of set difference operator and give an example to illustrate the same.

Use of set difference operator:

Allows finding tuples that are in one relation but are not in another relation.

Example: Find all customers of the bank who have an account but not a loan.

Relational Algebra Expression:

$\pi_{Customer\_name}$ (depositor) - $\pi_{customer\_name}$ (borrower)

23. What will be the output relation after performing the Cartesian product of borrower and loan relations? Consider two tuples for each relation.

Sample Relations:

a. Borrower Relation:

Cust_name	loan_no
Smith	L-17
James	L-11

b. Loan Relation:

loan_no	branch_name	amount
L-17	perryridge	1500
L-24	brighton	2000

c. Output Relation:

cust_name	Borrower.loan_no	Loan.loan_no	branch_name	amount
Smith	L-17	L-17	perryridge	1500
Smith	L-17	L-24	brighton	2000
James	L-11	L-17	perryridge	1500
James	L-11	L-24	brighton	2000

If n1 tuples are in borrower relation and n2 tuples in loan relation, then n1 * n2 tuples will be listed in the result relation of Cartesian product.

24. Explain the use of Assignment operator with an example.

Assignment operator is a convenient way to express complex queries. It is convenient at times to write a relational algebra expression in parts using assignment to a temporary relation variable.

Eg., We can write r ÷ s as

temp1 ← $\pi_{R-S}(r)$

temp2 ← $\pi_{R-S,S}$ (r)

temp3 ← $\pi_{R-S}(\pi_{R-S}(r)$ x s)

result ← temp1 – temp3 – temp2

temp relation does not result in any relation being displayed to user.

25. What are the operations available in the modification of database?

a. Insert to insert new record.

b. Deletion to delete the existing records

c. Update to modify the records.

26. What is a view?

Any relation that is not the part of the logical model but is made visible to a user as virtual relation is called a view. This mainly for projecting the needed record sets to the user for the security reasons. The syntax is CREATE VIEW view name as query.

27. What is the basic structure of SQL expression of retrieval?

SELECT clause corresponds to projection operation. FROM clause corresponds to the Cartesian-product operation of the relational algebra. WHERE clause corresponds to the selection predicate of the relational algebra.

28. Give the various operations in DDL.

       Create, alter and drop the objects.

29. Give the various operations in DML.

       Insert, select, delete and update the record values.

30. Give the various types of integrity constraints.

       Keys such as primary key, unique key, foreign key, composite key. Default, check and not null. These forms ensure the consistency and non-redundancy of data in the tables.

31. List down the keys you can set in DBMS.

       Super key is a set of one or more attributes that taken collectively, allows us to identify uniquely an entity in the entity set. Candidate keys are the minimal subset super keys. Primary key is one, which distinctly identifies the entity sets values.

32. Short note on order by clause.

       The order by clause causes the tuples in the result of a query to appear in sorted order. By default order by clause lists items in ascending order. We may specify desc for descending order or asc for ascending order.

   select * from loan order by amount desc, loan _number asc.

33. List down the set operations
    a. Union
    b. Intersect
    c. Except

34. List down the aggregate functions
    a. Average: avg
    b. Maximum: max
    c. Minimum: min
    d. Total: sum
    e. Count: count

35. Write the query for "find the average account balance at the Perryridge branch"

   Select avg (balance) from account where branch_name="Perryridge".

36. Write one example using not in operator.

   "Find all customers who do have a loan at the bank, but do not have an account at the bank"

       Select distinct customer_name from borrower where customer_name not in (select customer_name from depositor)

37. What are the basic types of forms?

     Tabular forms

     Single-row or column forms

     Sub-forms

     Switchboard forms.

38. Write the types of report?

     Tabular, label, groups or subroutine.

39. What is stored procedure?

       A stored procedure is a program that is executed through a single SQL statement that can be locally executed and completed within the process space of the database server.

40. What are the form creation tools available in creating oracle forms?
    a. Forms Designer
    b. Forms Generator
    c. Run Form

41. Name some common controls on forms
    a. Label control
    b. Text boxes
    c. Command button
    d. Toggle buttons

e. Check buttons
f. Option buttons
g. Combo box
h. List box
i. OLE container

42. What is driver?

A driver is a software program that translates the ODBC or JDBC calls into DBMS specific calls.

43. What are the methods available in the Driver manager class?

register Driver
deregister Driver
get Driver

44. What is trigger?

A trigger is a statement that is executed automatically by the system as a side effect of a modification to the database.

45. What are the three parts available in a trigger description?

Event
Condition
Action

46. What are the two parts of database functionality?

Front end, back end.

47. Write the categories of the server system?

Transaction server, data server.

48. What is decision support system?

Decision support system (DSS) also known as EIS (executive information systems) support an organization's leading decision makers with higher level data for complex and important decisions.

49. Name some partitioning techniques.

Round Robin, Hash partitioning, Range partitioning.

50. What are the classifications of the skew?

a. Attribute_value skew
b. Partition skew

51. What is Inter query parallelism?

In Inter query parallelism, different queries or transactions execute in parallel with one another. This form of parallelism can increase transaction throughput.

52. What is Intra query parallelism?

Intra query parallelism refers to the execution of a single query in parallel on multiple processors and disks.

53. Execution of a single query can be parallelized in two ways. What are they?

a. Intra operation Parallelism
b. Interoperation Parallelism

54. What is data parallism?

Execution of the same operation in parallel on different sets of data is called data parallism.

55. What are the two forms of interoperation parallelism?

a. Pipelined parallelism
b. Independent parallelism

56. What is independent parallelism?

Operations in a query expression that do not depend on one another can be executed in parallel. This form of parallelism is called independent parallelism.

57. Write some approaches for storing the relation in distributed data storage?

a. Replication

        b. Fragmentation

        c. Replication and fragmentation

58. Write the types of fragmentation?

        a. Horizontal fragmentation, vertical fragmentation, mixed fragmentation.

59. Short notes on Horizontal fragmentation.

The relation r is partitioned into a number of subsets r1,r2,…rn. Each tuple of relation r must belong to at least one of the fragments, so that the original relation can be reconstructed, if needed.

To reconstruct fragment ri as follows:

        i. $r_i = \sigma p_i (r)$

where r is a global relation, p_i is a predicate.

60. Short notes on vertical fragmentation.

In its simplest form, vertical fragmentation is the same as decomposition. Each fragment ri of r is defined by

    $r_i = \pi_{R_i} (r)$

we can reconstruct the relation r from the fragments by taking natural join

    a. $r = r1 \bowtie r2 \bowtie r3 \bowtie r4 \ldots \bowtie rn$

61. Write the issues of network transparency?

        a. Naming of data items

        b. Replication of data items

        c. Fragmentation of data items

        d. Location of fragments and replicas

62. Write the two disadvantages of naming of data items?

First, poor performance. Second, if the name server crashes, it may not be possible for any site in the distributed system to continue to run.

63. What is a global and local transaction?

The local transaction is those that access and update data in only one local database; the global transaction is those that access and update data in several local database.

64. Write the subsystems of each site?

The transaction manager, the transaction coordinator

65. What are the failure types of a system?

        Failure of a site

        Loss of messages

        Failure of a communication link

        Network partition

66. Name some network topology.

        Fully connected network

        Partially connected network

        Tree structured network

        Star network

        Ring network

67. What are the different ways for handling the failure?

        a. Retransmission of a message.

        b. Reconfigure the system.

68. What is backup coordinator?

A backup coordinator is a site that, in addition to other tasks, maintains enough information locally to allow it to assume the role of coordinator with minimal disruption to the distributed system. The main difference between the coordinator and its backup is that the backup does not take any action that affects other sites.

69. What is multidatabase system?

Manipulation of information located in a heterogeneous database requires an additional software layer on top of existing database system. This software layer is called multidatabase system.

70. What are the two types of transaction in a multidatabase system?

Local transaction, Global transaction.

71. What is valid time and transaction time in temporal databases?

The most natural interpretation is that the associated time is the time that the event occurred, or the period during which the fact was considered to be true in the read world. If this interpretation is used, the associated time is often referred to as the valid time. A temporal database using this interpretation is called a valid time database. The different interpretation is used, that is, it is the value of the system clock when the information is valid in the system. In this case, the associated time is called transaction time. A temporal database using this interpretation is called a transaction time database.

72. List the types of spatial queries?

Range query, nearest neighbour query, spatial joins or overlays.

73. List some types of multimedia data.

Text, graphics, animations, video, audio.

74. Name the two types of database accessible through web?

Access using CGI scripts, access using JDBC.

75. What is a type constructor?

The structuring operations needed for defining the structure of the state of object are called type constructors. They define basic structuring operations that can be combined to form complex object structures.

76. What are the goals of OODB?

The goals of OODB are:

i)   To maintain direct correspondence between real world and database object, so that the object will not loose its integrity and identity.

ii)   The objects, which are complex, need not be scattered in the database, and hence to create complex object structures.

77. What are persistent objects and transient objects?

Objects in OOPL exist only during program execution and are hence called transient objects.

Objects in OODB can be extended, so that they can exist in permanent storage even after the program termination. They are called persistent objects.

78. What is OID?

OO databases provide a unique system generated object-identifier (OID) for each object. Its value is not visible to the external user and they are used to create and manage inter-object references.

79. What are the characteristics an OID possess? (What are the properties of OID?)

The characteristics of OID are:

iii)   Immutable – They do not change.

iv)   An OID can be used only once.

80. What are the two components of object?

The two components of object are:

1) State(Value) and 2) Behavior(Operation).

81. Write the formal structure of an object.

An object is normally represented as a triple O = (i, c, v)

i-   The unique object identifier.

c- The type constructor.
v-   The object state or current value.

82. What are the six type constructors?
    The six type constructors are : Atom, Tuple, Array, List, Bag and Set.
  Array, List, Bag and Set are called Collection types or Bulk types.

83. Explain each type constructor.
  Atom – If c is atom, the value v is atomic and it is supported by the system.
  Set – The IODs for the set of objects are of same type. This does not allow duplicates.
  Array - The IODs for the set of objects are single dimensional array of object identifier.
  List – The list is similar to set, except that the OIDs are ordered.
  Bag – Bag is also called multiset. This can also contain duplicate elements.

84. What are identical objects and equal objects?
    Two objects are said to have equal states, if their states at the atomic levels are the same, but the values are reached through distinct objects.
    Two objects are said to be in identical states, if the objects are identical even though the objects themselves are not as they have distinct OIDs.

85. What is encapsulation?
  Encapsulation is also called information hiding, which is related to the concepts of abstract data types. This defines the behavior of a types of object based on the operation it has to perform. The internal structure of object is hidden and the object is accessible only through the predefined operations.

86. How can an operation be defined in encapsulation?
  The operation defined in encapsulation concept has two parts:
    v) Signature or interface of the operation – It specifies the operation name and arguments.
    vi) Method or Body of the operation – It specifies the implementation of operation.

87. What are hidden and visible attributes?
  Visible attributes are the attributes that may be directly accessed for reading by external operators or by high-level query language. Hidden attributes are the attributes that are completely encapsulated and can be only applied through predefined operations.

88. What is object constructor, destructor and object modifier?
    Object constructors are the operations, which are used to create new object. Object destructors are used to destroy objects. The object modifiers are the operations declared to modify various attributes of an object.

89. What are the methods for creating persistent objects?
    Persistent objects are the objects stored in databases that persist even after the program termination. The techniques to create persistence objects are: Naming and reachability.
    Naming mechanism involves giving an object a unique persistent name through which it can be retrieved with this and other programs. The named persistent objects are the entry points to the database.
    The reachability mechanism works by making the object reachable from some other persistent object. Making the object to be referenced by another persistent object B can make an object A persistent.

90. What is inheritance?

Inheritance is the concept of OO systems, that permits specification of new types or classes that inherit their structure and operations from previously defines types or classes.

91. What is persistent collection?

If we first create a names persistent object N, whose state is a set or list of objects of some class C, we can make the objects of C persistent by adding them to the set or list, and thus making it reachable from N. This N defines the persistent collection of objects of C.

92. What is sub types and super type?

The subtype is the type, which must be formed from already existing type by inheriting some of its features. The super type is the type from which the functions are inherited.

93. What is operator polymorphism ? (Operator overloading).

It refers to the operation's ability to be applied to different types of objects. The operation's name may refer to each distinct implementation, depending on the type of objects it is applied to.

94. What is late binding?

If the type of object to which the function is applied is not known until runtime and in this case the function must check the type of object and then invoke the appropriate method. This is called late binding.

95. What is early binding?

If the types of object for invoking the function is known at the compile time itself it is said to be early binding.

96. What is persistent collection and transient collection?

A persistent collection holds the collection of objects that is stored permanently in the database and hence can be accessed and shared by multiple programs.

A transient collection exists temporarily during execution of program but is not kept when the program terminates.

97. What are the two types of complex objects?

The two types of complex objects are:
   vii)   Structured complex object
   viii)   Unstructured complex object.

98. What is a structured complex object?

Structured complex object is made up of components that can be defined by applying available type constructor recursively.

99. What is unstructured complex object?

Unstructured complex objects are the complex objects that typically require a large amount of storage, such as a data type that requires a large amount of storage.

100.   What is ownership semantics?

Ownership semantics applies when the sub-objects of complex object are encapsulated within the complex object and are hence considered as part of complex object. This is also referred to as "is-part-of" or "is-component-of" relationship.

101. What is reference semantics?

Reference semantics are applied when the components of the complex objects are themselves independent objects but may be referenced from the complex object.

102. What is complex object assembly?

Storing the components of structured complex object on the same disk is called complex object assembly.

103. Define multiple inheritance and selective inheritance.

Multiple inheritance in a type hierarchy occurs when a certain subtype T is a subtype of two types and hence inherits the functions of both supertypes.

Selective inheritance occurs when a subtype inherits only some of the functions of supertype.

104. Define Data mining

Data mining refers to the mining or discovery of new information in terms of patterns or rules from vast amount of data.

105. Write the relationship between data warehouse and data mining.

i) Data warehouse is aggregate and summarized collection of data that makes data mining efficient.

ii) Data warehouse supports decision-making

iii) Data mining helps to extract meaningful new patterns

iv) Data mining can be applied to operational databases.

106. What are the different phases in Knowledge discovery?

i) Data Selection – Selecting data about specific item or category

ii) Data cleansing – Correcting invalid data or eliminating records

iii) Enrichment – Enhancing data with additional sources of information

iv) Data transformation and Encoding – Reducing amount of data by generalization

v) Data mining – Techniques to mine different rules and patterns

vi) Reporting and Display of discovered information – Displaying result as listings, graphical outputs, summary tables or visualizations in a user understandable manner

107. What are the results of data mining? Give examples for each.

i) Association rules – Whenever the customer buys a video equipment he is likely to buy electronic gadget.

ii) Sequential Patterns – If a customer buys camera, with in three months he will buy photographic accessories.

iii) Classification trees – Customers of a supermarket can be classified based on the frequency of visits.

108. What are the goals of Data mining?

i) Prediction – how certain attributes will behave in future.

ii) Identification - To identify existence of certain item, event or activity.

iii) Classification – Partitioning data in to different classes, based on combination of parameters.

iv) Optimization – Optimizing use of limited resources such as time, space, money and materials to maximize sales, profit.

109. What is knowledge? What are its types?

Knowledge is the degree of intelligence. It can be classified as inductive knowledge and deductive knowledge.

110. What are the ways to represent knowledge extracted during data mining?
   i) Association rules
   ii) Classification hierarchies
   iii) Sequential patterns
   iv) Patterns with in time series
   v) Categorization and Segmentation

111. What is Association rules?
   Association rules is the result of data mining. Association rule is of the form X=>Y. If customer buys X, he/She is likely to buy Y.

112. What is Support/Prevalence?
   Support for the rule LHS=>RHS is the percentage of transactions that hold all of the items in the union of the set LHS U RHS.

113. What is Confidence or Strength?
   For the association rule LHS=>RHS, consider all the transactions that include LHS. Confidence is the percentage of these transactions that include RHS.

114. Why mining of association rules are more complicated?
   i) The relationship between item sets is very large and the volume of transaction is very high as well.
   ii) Transactions show variations based on the factors like geographic locations, seasons and make sampling difficult.
   iii) Item classifications exist along multiple dimensions
   iv) Quality of data is variable.

115. What are the different approaches to data mining problems?
   i) Discovery of sequential patterns
   ii) Discovery of patterns with in time series
   iii) Discovery of classification rules
   iv) Regression
   v) Neural networks
   vi) Genetic algorithms
   vi) Clustering and segmentation

116. What is regression?
   Regression is a similar approach to classification. If the classification rules is regarded as a function over the variables that maps these variables into a target class variable, the rule is called regression rule.

117. What is a regression function?
   The function P=f(test1,test2,…testn) for mapping the patient test information into the target variable P, the probability of survival of patient in the LAB_TESTS(patientid,test1,test2,…testn) is called the regression function.

118. What is neural network? What are its two types?
   Neural network is the technique developed from artificial intelligence. It uses generalized regression and provides an iterative method to carry it out.
   The two categories of neural network are:

i) Supervised learning – Adaptive methods taken to reduce the output error.
ii) Unsupervised learning – They develop internal representations with out sample outputs.

119. What are genetic algorithms? What are its drawbacks in data mining?
Genetic algorithms are randomized search procedures, capable of adaptive and robust search over a wide range of search space technologies. They construct genetic algorithms from four letter genetic alphabet {A,C,T,G}
Drawbacks are: Large over production of individual solutions, Random character of searching process, High demand on computer processing.

120. What is clustering?
Clustering is the data mining technique, that is directed towards goals of identification and classification. It identifies the finite set of categories or clusters to which each data object can be mapped.

121. What are the applications of data mining?
  i)   Marketing
  ii)   Finance
  iii)   Manufacturing
  iv)   Health care

122. Define Data Warehouse.
Data warehouse is a subject-oriented, integrated, non-volatile, time-variant collection of data in support of management's decisions. (OR) Data warehouse is a collection of decision support technologies, aimed at enabling the knowledge worker to make better and faster decisions. (OR) Data warehouse is a repository of information gathered from multiple sources, stored under a unified schema, at a single site.

123. What is the goal of data warehouse?
The goal of data warehouse is to provide access to data for complex analysis, knowledge discovery and decision-making. They also support OLAP, DSS and Data mining.

124. What are the Applications of Data warehousing?
i) OLAP (Online Analytical Processing)
ii) DSS (Decision Support System)/EIS (Executive Information Systems)
iii) Data Mining

125. Briefly explain about the Applications of Data warehousing.
OLAP- It has distributed computing capabilities, that require more storage and processing power and help in analysis of Complex Data.
DSS-They support Organization's decision makers with higher level data for complex and more important decisions.
Data mining- Used for Knowledge discovery, process of searching data for new knowledge.

126. What are the differences between relational databases and data warehouse?
  i) Relational databases support insertions and updates, while Data warehouse support extraction, Processing, Analysis and decision-making.
  ii) Relational database data may be distributed, while data warehouse is integrated data from multiple sources.

127. What is online transaction processing?

OLTP supports transactions in distributed environment. It supports insertions, updates, deletions and information query requirements.

128. What are the differences between transactional databases and data warehouse?
    i) Transactional database provides access to disjoint and heterogeneous databases.
      Data warehouse store integrated data from multiple
                  sources in multidimensional model
    ii) Transactional database support storage of historical data.
      Data warehouse support time-series and trend analysis with historical
      data.
    iii) Transactional databases are volatile
      Data warehouses are non-volatile.
    iv) Transactional database units are records.
      Data warehouse information is coarse grained, and subjected to
      incremental refreshing.

129. What are the distinctive characteristics of data warehouse?
    i) Multidimensional conceptual view
    ii) Generic Dimensionality
    iii) Unlimited dimensions and aggregation levels.
    iv) Unrestricted cross dimensional operations
    v) Dynamic Sparse matrix handling
    vi) Client-server architecture
    vii) Multi-user support
    viii) Accessibility
    ix) Transparency
    x) Intuitive data manipulation
    xi) Consistent reporting performance
    xii) Flexible reporting

130. What are the types of data warehouse?
    i) Enterprise-wide data warehouse – For huge projects with large investment and resources
    ii) Virtual data warehouse – Provide views of operational databases materialized for efficient access.
    iii) Data mart – Data from any organization or department.

131. What is data model? What are the types?
Data model is the structure in which the data is stored.
Two dimensional data models are spreadsheets. Three dimensional matrices are data cubes and data model with more than three dimensions are hypercubes.
(Include the diagrammatic representation too)

132. Define Pivoting or Rotation.
Changing from one-dimensional hierarchy to another is called Pivoting.

133. What are the two types of views in multidimensional model?
    i) Roll-up display – Move sup the hierarchy, by grouping into larger units along a dimension. This is a coarser grained view, which increases generalization.
    ii) Drill-down Display – This furnishes a finer-grained view.
    (Include the diagrammatic representation too)

134. What are the two types of tables in multidimensional model?
      i)Dimension table – Consists of tuples of attributes of the dimension.
      ii)Fact table – Consists of tuples, one per recorded format. It has variables and pointers to dimension table.

135. What are the two types of multidimensional schemas?
      i)Star schema – It consists of fact table with single table for each dimension
      ii)Snowflake schema – The dimension tables from Star schema are organized hierarchically by normalizing them.
      (Include the diagrammatic representation too)

136. Define Fact constellation.
Fact constellation is the set of tables that share sae dimension tables. They limit the possible queries for the data warehouse.
      (Include the diagrammatic representation too)

137. What is indexing? What are the two types of indexing?
Indexing is the technique to improve high performance access. Two types are,
i) Bit map indexing – It constructs a bit vector for each value in domain being indexed. A 1-bit will be placed in jth position of corresponding bit vector, if jth row contains the domain being indexed.
ii) Join indexing – This is used to maintain the relationships between primary key and foreign key.

138. What are the issues to be considered while building data warehouse?
i) When and how to gather data
ii) What schema to use
iii) How to propagate update
iv) What data to summarize

139. What are the steps to acquire data for data warehouse?
i) Data extraction
ii) Consistency maintenance
iii) Cleaning and Back flushing
iv) Fitting
v) Loading

140. What is back flushing?
The process of returning the cleaned data to the source is called back flushing.

141. What is fitting?
Conversion of data from relational, object-oriented to multidimensional model for transferring the data from database to multidimensional model is called fitting.

142. What are the steps in data storage into data warehouse?
i) Fitting data into data model
ii) Creating and maintaining required data structure
iii) Creating and maintaining access paths
iv) Providing time-variant data
v) Support updation of warehouse data
vi) Refreshing data

vii) Purging/ Cleaning of data

143. What are the design considerations of data warehouse based on environment?
i) Usage Projections
ii) The fit of the data model
iii) Characteristics of available resources
iv) Design of metadata component
v) Modular component design
vi) Design for manageability and change
vii) Considerations of distributed and parallel architecture

144. What is a metadata repository?
The information stored in the Catalog to store the primary structure of data base is called metadata. Metadata repository are the key components of data warehouse, containing Technical metadata and Business metadata.
Technical metadata – Contains details of acquisition processing, storage structures, data descriptions, warehouse operations, maintenance and access support functionality.
Business metadata – Contains business rules and organizational details.

145. What are the two types of distributed architectures of data warehouse?
i)Distributed warehouse – Distributing metadata by replication, partitioning, communication and consistency maintenance.
ii)Federated data warehouse – Decentralized autonomous data warehouse with its own metadata repository. The small units of it are data mart.

146. What are the benefits of distributed data warehouse?
To improve warehouse performance such as load balancing, scalability of performance and high availability.

147. What are the functionalities of data warehouse?
i) Roll-up – Increasing generalization
ii) Drill-down – Increasing levels of details
iii) Pivot - Rotation or Cross-tabulation is performed
iv) Slice and Dice – Projection operations on the dimensions
v) Sorting – Arranging data by ordinal value
vi) Selection – Choosing data by value or range
vii) Derived or Computed attribute – Computing attributes by operations on stored and derived values.

148. What are the tools and techniques for query processing?
i) Query transformation
ii) Index intersection and union
iii) Special ROLAP and MOLAP
iv) SQL Extensions
v) Advanced Join methods
vi) Intelligent scanning

149. Discuss on the common features and difference between data warehouse and views?

Common: Views and data warehouses have read-only extracts from databases, and they both are subject oriented.

Difference:

Storage: DW – Persistent storage

Views – Materialized storage on demand

Model: DW – Multidimensional model

Views – Relational model

Indexing: DW – Can be indexed to optimize performance

Views – Cannot be indexed

Functionality: DW – Support Roll-up, Drill-down, Pivot, Slice and dice etc

Views – Cannot support functionalities like DW.

Data: DW – Provide integrated and temporal data contained in more than one database.

Views – Views are extract of database

150. What are the difficulties in Implementing Data warehouse?

i) Construction, Administration and Quality control are difficult tasks.

ii) Building enterprise-wide data warehouse is time consuming process.

iii) Maintenance of consistency while getting data from different resources is very difficult as the naming of domains and identification will vary.

iv) Usage projections need to be revised periodically.

v) Team of experts with overlapping areas of expertise is needed.

vi) Effective data warehouse management with technical skills, careful coordination and effective leadership is needed.

16 Marks Questions

1. Briefly explain about fundamental, additional operations in relational algebra with example?

Six Fundamental operators:
- Selection (σ)
- Projection (π)
- Union ($\cup$)
- Set Difference (-)
- Cartesian Product (X)
- Rename (ρ)

Four Additional operators:
- Set Intersection ($\cap$)
- Natural Join ($\bowtie$)
- Division ($\div$)
- explain in detail with example.

2. Briefly explain about the modification of the database?

Explain-deletion, insertion and updation with example.

3. Explain about basic structure of SQL with example.

Explain select clause, form clause and where clause with example.

4. Write the Query using the following.

Order by clause, in, not in, exist, as clause.

-Write the query using the above.

5. Briefly explain about decision support system.

-Definition for decision support system, give example, about data warehouse.

6. Explain about forms and reports.
   Forms- Uses of forms, guideline to design form, types of forms, creating oracle forms.
   Report-Uses of report effective design of report, types of report.
7. Explain about ODBC and JDBC.
   Explain-ODBC, JDBC classes and interfaces, JDBC driver management, connections and executing SQL statements.
8. Explain about stored procedure and trigger.
   Definitions for stored procedure and explain with example.
   Definition for trigger and explain with example.
9. Briefly explain about set operations and aggregate functions.
   Set operations
   - Union
   - Intersect
   - Except
   Aggregate functions
   - Average: avg
   - Maximum: max
   - Minimum: min
   - Total: sum
   - Count: count
10. List the partitioning techniques and explain them.
    Explain - Round Robin, Hash partitioning, Range partitioning.
11. Briefly explain about inter query parallelism and intra query parallelism.
    Explain - Inter query parallelism and Intra query parallelism.
12. Briefly explain about intra operation parallelism and interoperation parallelism.
    Explain
    -Intraoperation parallism-parallel sort, Range partitioning sort, parall external merge sort
    -Interoperation parallism - pipelined parallism, independent parallism.
13. Briefly explain about distributed database.
    Explain- Distributed data storage-Replication, fragmentation, Replication and fradmentation.
14. Explain about commit protocol.
    Two phase commit.-phase1,phase2.
    Handling of failures-Failure of participating site, failure of the coordinator, network partition.
15. Explain about coordinator selection and concurrency control.
    Coordinator selection- backup coordinator, bully algorithm.
    Concurrency control- locking protocol
16. Explain about deadlock handling.
    Define deadlock, deadlock characterization, local, global weight of graph.
17. Explain about deductive database.
    Explain- deductive database, prolog/datalog notation, safety of programs.
18. Explain about Temporal database.
    Explain-definition, examples, valid time, transaction time, bitemporal database, implementation consideration, attribute versioning.
19. Explain about web database.
    Explain- definition, HTML, uniform resource locator, access using CGI script, access using JDBC, oracle webserver.
20. Explain about spatial and multimedia database.
    Spatial-definition, example, spatial queries.
    Multimedia database-definition, example,
    types of multimedia data-Text, graphics, animations, video, audio. Hypermedia, applications.

21. Write about type constructors and how complex objects can be formed with type constructors with example. (16).

   Hints : Define type constructor – Atom , Tuple, Array, Set, List, Bag – Explain each with their structure and difference – Collection/ Bulk types – Eg: Complex object struture – Department-Draw the graph.

22. What are identical and equal objects? (4)

   Hints: Give example – Define identical objects – Equal objects.

23. What is the difference between DBMS and OODB in complex object representation?

24. Explain Complex objects in detail. (16)

   Hints: Types of complex object – Structured complex object : Eg, Ownership semantics, reference semantics, clustering – Unstructured complex object : Eg-BLOBs, large texts, storage and retrieval in DBMS and OODB.

25. What is versioning and configuration? What is the difference between them? (8).
Hints: Version- Version graph – Configuration – Concurrent engineering - Difference.
26. Explain in detail about data warehouse and its implementation.

   Hints:
   Define Data warehouse- Draw the conceptual structure of data warehouse and explain-Issues to be addressed in building a data warehouse-Steps to acquire data warehouse-Steps in data storage-Design considerations for data warehouse building-Meta data repository-Distributed architectures of data warehouse-Difficulties in building Data warehouse
27. Explain about data mining in detail.

   Hints: Define Data mining-Relation between data warehouse and data mining-Six phases in knowledge discovery-Goals of data mining-How data mining results are represented with example of each.

28. Explain in detail about Association rules of data mining.

   Hints: Form of association rule-Support-Confidence-Transaction example-Basic algorithm for association rule-Partitioning algorithm-Association rule among hierarchies-Negative associations-Why association rule mining is more complicated?

29. What are the approaches to data mining problems?

   Hints: Discovery of sequential patterns-Discovery of patterns with in time series-Discovery of classification rules-Regression-Neural networks-Genetic algorithms-Clustering and segmentation

30. Explain about Clustering.
   Hints: Define Clustering – Types of clustering – K-means & K-medoids clustering algorithm-Applications of clustering.