DBMS Questions

1. What is a database?
Ans: A database is a collection of related data. A database is a logically coherent collection of data with some inherent meaning.

2. What is DBMS?
Ans: Database Management system is a collection of programs that enables user to create and maintain a database.

3. What is a Catalog?
Ans: A catalog is a table that contains the information such as structure of each file, the type and storage format of each data item and various constraints on the data. The information stored in the catalog is called Metadata. Whenever a request is made to access a particular data, the DBMS s/w refers to the catalog to determine the structure of the file.

4. How is the data stored in a database?
Ans: Data is stored in the form of tables.

5. What is a table?
Ans: A table is a collection of records. Record is also known as a Row.

6. What is a column/field?
Ans: It is known as an attribute which is a property of a table.

7. What is SQL?
Ans: Structured Query language.

8. What are the applications of databases?
Ans: Airlines, banking, universities, credit card transactions, tele communications, Finance, Sales etc.

9. In E-R Model what does the term E and R mean?
Ans: E means Entity and **R** means Relation.

10. In E-R model attributes are represented using Oval symbol.

11. In E-R model relations are represented using Rhombus symbol.

12. FPS stands for?
Ans: File Processing System.

13. What is data Redundancy?
Ans: Duplication of data is called data redundancy.

14. What is real time database technology?
Ans: These are all the techniques used in controlling industrial and manufacturing processes.

15. What is program-data independence?
Ans: Unlike in the traditional file sys. the structure of the data files is stored in the DBMS catalog separately from the access programs . This property is called program-data independence.

16. What is ORDBMS?
Ans: Object oriented RDBMS is a relational DBMS in which every thing is treated as objects. User can define operations on data as a part of the database definition.

17. What is program-operation independence?
Ans: An operation is specified in two parts.
1. Interface (operation name and data types of its arguments).
2. Implementation (the code part)
The implementation part can be changed without affecting the interface. This is called program-operation independence.

18. What is a view?
Ans: A view may be a subset of the database or it may contain virtual data that is derived from the database files but is not explicitly stored.

19. What are the different database models?
Ans: Flat model, Hierarchical models, Network model, Relational model and Dimensional model.

20. What is OLTP?
Ans: Online transaction processing is an application that involve multiple database accesses from different parts of the world. OLTP needs a multi-user DBMS s/w to ensure that concurrent transactions operate correctly.

21. What is the job of DBA?
Ans: A database administrator is a person or a group responsible for authorizing access to the database, for coordinating and monitoring its use, and for acquiring s/w and h/w resources as needed.

22. Who are db designer?
Ans: Data base designers are responsible for identifying the data to be stored in the database and for choosing appropriate structure to represent and store this data.

23. What are different types of end users?
Ans:
1. Casual end-users
2. Naive or parametric end users
3. Sophisticated end users
4. Stand alone users.

24. What are the advantages of using a dbms?
Ans:
1. Controlling redundancy.
2. Restricting unauthorized access.
3. Providing persistent storage for program objects and data structures.

4. Providing multi-user interfaces.
5. Representing complex relationships among data.
6. Enforcing integrity constraints.
7. Providing backups and recovery.

25. **What are the disadvantages of using a dbms?**
Ans:
1. High initial investments in h/w, s/w, and training.
2. Generality that a DBMS provides for defining and processing data.
3. Overhead for providing security, concurrency control, recovery, and integrity functions.

26. What is a data model?
Ans: It is a collection of concepts that can be used to describe the structure of a database. It provides necessary means to achieve this abstraction. By structure of a database we mean the data types, relations, and constraints that should hold on the data.

27. What are different categories of data models?
Ans:
1**. High-level** or conceptual data models.
2. **Representational data models**.
3. **Low-level** or physical data models.
High level data models provide the concepts that are close to the way many users perceive data.
Representational data models are provide concepts that provide the concepts that may be understood by end users but that are not too far removed from organization of data in the database.
Physical data models describe the details of how data is stored in the computers.

28. **What is schema**?
Ans: The description of a data base is called the database schema, which is specified during database design and is not expected to change frequently. A displayed schema is called schema diagram.

29. What is Instance of Database?
Ans: The collection of information stored in the database at a particular movement is called the **Instance** of the Database.

30. What are types of schema?
Ans:
1. Internal schema defines physical storage structures.
2. Conceptual schema integrates external schema.
3. External schemas or user views.

31. What is Data independency?
Ans: Data independency is defined as the capacity to change the conceptual schema without having to change the schema at the next higher level.

32. What are the types of Data Independency?
Ans: We can define two types of data independence:
1. Logical data independence (LDI).
2. Physical data independence (PDI).

LDI is the capacity to change the conceptual schema without having to change external schemas or application programs.
PDI is the capacity to change the internal schema without having to change conceptual (or external) schemas.

33. What are different DBMS languages?
Ans:
1. DDL (Data definition language)
2. DML (Data manipulation language)
3. DRL (Data Retrieval Language)
4. DCL (Data Control Language)
5. TCL (Transaction control language)

34. What are different types of DBMS?
Ans:
1. DBMS
2. RDBMS (Relational)
3. ORDBMS (Object Relational)
4. DDBMS (Distributed)
5. FDBMS (Federated)
6. HDDBMS (Homogeneous)
7. HDBMS (Hierarchical)
8. NDBMS (Networked)

35. What is an entity?
Ans: An entity is a thing in the real world with an independent existence.

36. What are attributes?
Ans: These are the particular properties that describe an entity.

37. What are diff. types of attributes?
Ans:
1. Composite Vs simple attributes.
2. Single valued Vs multi-valued attributes.
3. Stored Vs derived attribute.
4. Null valued attributes.
5. Complex attributes.

38. What is difference between entity set and entity type?
Ans:
Entity type: An entity type defines a collection of entities that have the same attributes.
Entity set: the collection of all entities of a particular entity type in the database at any point of time is called Entity set.

39. What is domain value or value set of an attribute?
Ans: It is the set of values that may be assigned to that attribute for each individual entities

40. What is degree of a relationship?
Ans: The no of entities participating in that relation.

41. What is recursive relationship?
Ans: It is the relationship where both the participating entities belong to same entity type

42. What is cardinality ratio in DBMS?
Ans: The ration between total Participation and partial participation is called as Cardinality ration.

43. What are relationship constraints?
Ans:
1. Cardinality ratio.
2. Participation constraints.

44. What is a Participation constraint?
Ans: The participation constraint specifies whether the existence of an entity depends on its being related to another entity via the relationship type. This is of two types:
1. Total participation.
2. Partial participation.

45. What is a weak entity types?
Ans: The entity types that do not have key attributes of their own are called weak entity types.

46. What is an ER Diagram?
Ans:
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.

47. What is meta data?
Ans: Data about the data is called meta data.

48. What is specialization?
Ans: It is the process of defining a set of subclasses of an entity type where each subclass contain all the attributes and relationships of the parent entity and may have additional attributes and relationships which are specific to itself.

49. What is generalization?
Ans: It is the process of finding common attributes and relations of a number of entities and defining a common super class for them.

50. What is a ternary relationship?
Ans: A relationship with a degree 3 is called a ternary relationship.

51. What is aggregation and association?
Ans: Aggregation is an abstraction concept for building composite objects from their component objects. The abstraction of association is used to associate objects from several independent classes.

52: What is RAID?
Ans: Redundant array of inexpensive (or independent) disks.

53. What is RAID Technology?
Ans: Redundant array of inexpensive (or independent) disks. The main goal of raid technology is to even out the widely different rates of performance improvement of disks against those in memory and microprocessor. Raid technology employs the technique of data striping to achieve higher transfer rates.

54. What is Hashing technique?
Ans: This is a primary file organization technique that provides very fast access to records on certain search conditions. The search condition must be an equality condition on a single field, called hash field of the file.
1. Internal hashing
2. External hashing
3. Extendible hashing
4. Linear hashing
5. Partitioned hashing.

55. What are different types of relational constraints?
Ans:
1. Domain constraints
2. Key constraints
3. Entity integrity constraints
4. Referential integrity constraints

Domain constraints specify that the value of each attribute must be an atomic value from the domain of the attributes.
Key constraints tell that no two tuples can have the same combination of values for all their attributes.
Entity integrity constraint states that no primary key value can be null.
Referential integrity constraints states that a tuple in one relation that refers to another relation must refer to an existing tuple in that relation it is specified between two relations and is used to maintain the consistency among tuples of the two relations.

56. What is a super key?
Ans: A **superkey** is defined as a set of attributes of a relation variable for which it holds that in all relations assigned to that variable, there are no two distinct tuples (rows) that have the same values for the attributes in this set.

57. What is a Candidate Key?
Ans: Candidate key - A candidate key is a field or combination of fields that can act as a primary key field for that table to uniquely identify each record in that table.

58. What is a Compound key?
Ans: Compound key (also called a composite key or concatenated key) is a key that consists of 2 or more attributes.

59. What is a primary key?
Ans: A primary key is a value that can be used to identify a unique row in a table.

60. What is a foreign key?
Ans: A key of a relation schema is called as a foreign key if it is the primary key of some other relation to which it is related to.

61. What is difference between a super key, a key, a candidate key and a primary key?
Ans: A super key specifies uniqueness constrain that no two distinct tuples in a state can have the same value for the super key. Every relation has at least one default super key. A key is a minimal super key or the subset of the super key which is obtained after removing redundancy. A relation schema may have more than one key. In this case each key is called a candidate key. One of the candidate key with minimum number of attributes is chosen as primary key.

62. What is a transaction?
Ans: A transaction is a logical unit of database processing that includes one or more database access operations.

63. What are the properties of transaction?
Ans:
1. Atomicity
2. Consistency preservation
3. Isolation
4. Durability (permanence)

64. What are the basic data base operations?
Ans:
1. Write_item(x)
2. Read_item(x)

65. What are the disadvantages of not controlling concurrency?
Ans:
1. Lost update problem
2. Temporary update (Dirty read) problem
3. Incorrect summary problem

66. What are serial, non serial schedules?
Ans: A schedule S is serial if, for every transaction T participating in the schedule, all the operations of T is executed consecutively in the schedule, otherwise, the schedule is called non-serial schedule.

67. What is result equivalent?
Ans: Two schedules are called result equivalent if they produce the same final state of the data base.

68. What are conflict equivalent schedules?
Ans: Two schedules are said to be conflict equivalent if the order of any two conflicting operations is the same in both schedules.

69. What is a conflict serializable schedule?
Ans: A schedule is called conflict serializable if it is conflict equivalent to some serial schedule.

70. What is view serializable?
Ans: A schedule is said to be view serializable if it is view equivalent with some serial schedule.

71. What are the various methods of controlling concurrency?
Ans:
1. Locking
2. Time stamp
Locking data item to prevent multiple transactions from accessing the item concurrently.
A time stamp is a unique identifier for each transaction, generated by the system.

72. What is a lock?
Ans: A lock is a variable associated with a data item that describes the status of the item with respect to the possible operations that can be applied to it.

73. What are various types of locking techniques?
Ans:
1. a binary lock
2. Shared/Exclusive lock
3. Two phase locking

74. What is a binary lock?
Ans: A binary lock can have two states or values:
1. locked (1)
2. Unlocked (0)
If locked it cannot be accessed by any other operations, else can be.

75. What is shared or exclusive lock?
Ans: It implements multiple-mode lock. Allowing multiple accesses for read operations but exclusive access for write operation.

76. Explain two phase locking?
Ans: All the locking operations must precede the first unlock operation in the transaction .It does have two phases:
1. expanding phase (Locks are issued)
2. Shrinking phase (Locks are released)

77. What is a deadlock?
Ans: Dead lock occurs when each transaction T in a set of two or more transactions is waiting for some item that is locked by some other transaction T' in the set. Hence each transaction is in a waiting queue, waiting for one of the other transactions to release the lock on them.

78. What are triggers?
Ans: Triggers are the PL/SQL blocks defining an action the database should take when some database related event occurs. Triggers may be used to supplement declarative referential integrity, to enforce complex business rules, or to audit changes to data.

79. What is a surrogate key?
Ans: A surrogate key is a system generated sequential number which acts as a primary key.

80. What are the disadvantages of FPS?
Ans:
- Data redundancy & inconsistency.
- Difficult in accessing data.
- Data isolation.
- Data integrity.
- Concurrent access is not possible.
- Security Problems.

81. What are the three levels of abstractions?
Ans:
Physical level: The lowest level of abstraction describes how data are stored.
Logical level: The next higher level of abstraction, describes what data are stored in database and what relationship among those data.
View level: The highest level of abstraction describes only part of entire database.

82. Define the "integrity rules"
Ans: There are two Integrity rules.
Entity Integrity: States that Primary key cannot have NULL value?
Referential Integrity: States that Foreign Key can be either a NULL value or should be Primary Key value of other relation.

83. What is Object Oriented model?
Ans: 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.

84. What is Relationship?
Ans: It is an association among two or more entities.

85. What is Relationship set?
Ans: The collection (or set) of similar relationships.

86. What is Relationship type?
Ans: Relationship type defines a set of associations or a relationship set among a given set of entity types.

87. What is DML?
Ans: This language that enable user to access or manipulate data as organized by appropriate data model.

88. What is DDL?
Ans: Data Definition language deals with the structure of the table.

89. What is VDL (View Definition Language)?
Ans:
It specifies user views and their mappings to the conceptual schema.

90. What is DML Compiler?
Ans: It translates DML statements in a query language into low-level instruction that the query evaluation engine can understand.

91. What is Query evaluation engine?
Ans: It executes low-level instruction generated by compiler.

92. What is DDL Interpreter?
Ans: It interprets DDL statements and records them in tables containing metadata.

93. What is Record-at-a-time?
Ans: The Low level or Procedural DML can specify and retrieve each record from a set of records. This retrieve of a record is said to be Record-at-a-time.

94. What is Set-at-a-time or Set-oriented?
Ans: The High level or Non-procedural DML can specify and retrieve many records in a single DML statement. This retrieve of a record is said to be Set-at-a-time or Set-oriented.

95. What is Relational Algebra?
Ans: It is procedural query language. It consists of a set of operations that take one or two relations as input and produce a new relation.

96. What is Relational Calculus?
Ans: It is an applied predicate calculus specifically tailored for relational databases proposed by E.F. Codd.

97. How does Tuple-oriented relational calculus differ from domain-oriented relational calculus
Ans: The tuple-oriented calculus uses tuple variables i.e., variable whose only permitted values are tuples of that relation. E.g. QUEL
The domain-oriented calculus has domain variables i.e., variables that range over the underlying domains instead of over relation. E.g. ILL, DEDUCE.

98. What is normalization?
Ans: It is a process of analyzing the given relation schemas based on their Functional Dependencies (FDs) and primary key to achieve the properties.
-Minimizing redundancy
- Minimizing insertion, deletion and update anomalies.

99. What is functional dependency?
Ans: A functional dependency (FD) is a constraint between two sets of attributes in a relation from a database.

100. What is Lossless join property?
Ans: It guarantees that the spurious tuple generation does not occur with respect to relation schemas after decomposition.

101. What does SQL stand for?
Ans: Structured Query Language.

101. What is SQL?
Ans: SQL (Structured Query Language) is a standard language for accessing databases.

102. Name a few database systems.
Ans: MySQL, SQL Server, Access, Oracle, Sybase, DB2 etc..

103. Which SQL statement is used to extract data from a database?
Ans: select statement.

104. Which SQL statement is used to update data in a database?
Ans: Update

105. Which SQL statement is used to delete data from a database?
Ans: delete.

106. How do you create a table? Show with an example.
Ans: create table student(name varchar2(15), id number(7), branch varchar2(10));

107. How do you insert records into a table? Show with an example.
Ans: insert into student values('smith' 100, 'IT');

108. What is truncate in DBMS?
Ans: It deletes all the records in the table. It results in an empty table.

109. What is drop?
Ans: It removes records and table.

110: What is delete statement?
Ans: The DELETE statement is used to delete rows in a table.

111. What is the difference between delete and drop?
Ans: After performing a **DELETE** operation you need to COMMIT or ROLLBACK the transaction to make the change permanent or to undo it.
The **DROP** command removes a table from the database. All the tables' rows, indexes and privileges will also be removed. No DML triggers will be fired. The operation cannot be rolled back.

112. What is ALTER statement used for?
Ans: It is used to modify a table.

113. What are the various DDL operations?
Ans: create, alter, truncate, drop

114. What are the various DML operations?
Ans: insert, delete, update

115. What is DRL?
Ans: DRL stands for **data retrieval language.**

116. What is DCL and what are its operations?
Ans: DCL stands for Data control language. Its operations are **grant** and **revoke**

117. What is Grant?
Ans: GRANT is a command used to provide access or privileges on the database objects to the users.

118. What is revoke?
Ans: The REVOKE command removes user access rights or privileges to the database objects.

119. What is TCL and what are its operations?
Ans: TCL stands fore Transaction Control Language. Its operations are **Savepoint, Rollback and Commit.**

120. What is a Savepoint statement?
Ans: The SAVEPOINT statement names and marks the current point in the processing of a transaction.

121. What is Commit statement?
Ans: The COMMIT statement makes permanent any changes made to the database during the current transaction.

122. What is Rollback Statement?
Ans: The ROLLBACK statement is the inverse of the COMMIT statement. It undoes some or all database changes made during the current transaction.

123. Write a query to show the details of all the employees whose salaries is in the range 20,000 to 50,000$.
Ans: select * from emp where sal >=20,000 and sal <= 50,000

124. Write a query to list the employees whose name start with j and ends with s.
Ans: select * from emp where ename like 'j%s'

125. What are SQL Aggregate functions?
Ans: SQL aggregate functions return a single value, calculated from values in a column.
Useful aggregate functions:
AVG() - Returns the average value
COUNT() - Returns the number of rows
FIRST() - Returns the first value
LAST() - Returns the last value
MAX() - Returns the largest value
MIN() - Returns the smallest value
SUM() - Returns the sum

126. What is groupby statement?
Ans: The GROUP BY statement is used in conjunction with the aggregate functions to group the result-set by one or more columns.

127. What is orderby statement?
Ans: The GROUP BY statement is used in conjunction with the aggregate functions to group the result-set by one or more columns.

128. What is a SQL join?
Ans: The JOIN keyword is used in an SQL statement to query data from two or more tables, based on a relationship between certain columns in these tables.

129. What are different SQL joins?
Ans:
JOIN: Return rows when there is at least one match in both tables
LEFT JOIN: Return all rows from the left table, even if there are no matches in the right table
RIGHT JOIN: Return all rows from the right table, even if there are no matches in the left table
FULL JOIN: Return rows when there is a match in one of the tables

130. What is inner join?
Ans: The INNER JOIN keyword return rows when there is at least one match in both tables.

131. What is a constraint? And what are the levels of constraints?
Ans: Constraint is the restriction on values that can be entered in a table. There are two levels of constraints. 1. Column level 2. Table level.

132: What are the types of constraints?
Ans: We have three types of constraints.

1. Domain integrity constraint.
2. Entity integrity constraint
3. Referential constraint

133. What is column level constraint?
Ans: The constraint that is based on a single column and is applicable to that column only. All types of constraints can be placed at column level.

134. What is table level constraint?
Ans: In table level constraint, the constraints can be placed on more than one column. All types of constraints can be placed at Table level **except NOT NULL.**

135. What is the need for Locks in DBMS?
Ans: Locking is needed to allow only one user at a time to access the table. Other users are made to wait until the lock is released.

136. What is a database trigger?
Ans: A **database trigger** is procedural code that is automatically executed in response to certain events on a particular table or view in a database. The trigger is mostly used for keeping the integrity of the information on the database.

137. What is a Cursor?
Ans: A cursor comprises a control structure for the successive traversal (and potential processing) of records in a result set. Database programmers use cursors for processing individual rows returned by the database system for a query.

138. What is database Normalization?
Ans: Database normalization is the process of removing redundant data from your tables in to improve storage efficiency, data integrity, and scalability.

139. What is 1st normal form?
Ans: First normal form (1NF) sets the very basic rules for an organized database:
Eliminate duplicative columns from the same table.
Create separate tables for each group of related data and identify each row with a unique column or set of columns (the primary key).

140. What is 2nd Normal form?
Ans: In order to be in Second Normal Form, a relation must first fulfill the requirements to be in First Normal Form. Additionally, each non-key attribute in the relation must be functionally dependent upon the primary key.

141. What is 3rd Normal form?
Ans: In order to be in Third Normal Form, a relation must first fulfill the requirements to be in Second Normal Form. Additionally, all attributes that are not dependent upon the primary key must be eliminated.

142. What Boyce-Codd Normal Form (BCNF)?
Ans: A relation is in Boyce-Codd Normal Form (BCNF) if every determinant is a candidate key.

143. What is 4th Normal form?
Ans: To be in Fourth Normal Form, a relation must first be in Boyce-Codd Normal Form. Additionally, a given relation may not contain more than one multi-valued dependency.

144. What is distributed database?
Ans: A distributed database is defined as collection of logically distributed database which are connected with each other through a network. A distributed database management system is used for managing distributed database. Each side has its own database and operating system.

145. What is centralized database?
Ans: A centralized database has all its data on one place. As it is totally different from distributed database which has data on different places. In centralized database as all the data reside on one place so problem of bottle-neck can occur, and data availability is not efficient as in distributed database.

146. What is Data fragmentation?
Ans: Data fragmentation occurs when a piece of data is not able to fit in (available) memory-slot as a whole and is broken up into many pieces (that are not close together) so that each piece is able to fit the available memory locations.

148. What is internal fragmentation?
Ans: Internal fragmentation occurs when storage is allocated without ever intending to use it. This space is wasted. It is often accepted in return for increased efficiency or simplicity.

149. What is external fragmentation?
Ans: External fragmentation refers to the division of free storage into small pieces over a period of time, due to an inefficient memory allocation algorithm, resulting in the lack of sufficient storage for another program because these small pieces are not contiguous.

150. What is concurrency control?
Ans: Concurrency control mechanism ensures that correct results for concurrent operations are generated, while getting those results as quickly as possible.

151. What is Two-Phase Commit?
Ans: Two-phase commit is mechanism that guarantees a distributed transaction either commits on all involved nodes or rolls back on all involved nodes to maintain data consistency across the global distributed database. It has two phases, a Prepare Phase and a Commit Phase.

152. What is data mining?
Ans: Data mining the process of extracting patterns from data.

153. What is data warehousing?
Ans: A data warehouse is a repository of an organization's electronically stored data. Data warehouses are designed to facilitate reporting and analysis.

154. Differentiate between Data Mining and Data warehousing.
Ans: Data warehousing is merely extracting data from different sources, cleaning the data and storing it in the warehouse. Where as data mining aims to examine or explore the data using queries.

155. What is data mart?
Ans: A data mart is a subset of an organizational data store, usually oriented to a specific purpose or major data subject, which may be distributed to support business needs.

156. Explain the storage models of OLAP.
Ans: MOLAP Multidimensional Online Analytical processing - In MOLAP data is stored in form of multidimensional cubes and not in relational databases.

157. What is a Fact?
Ans: In data warehousing a fact is a value or measurement, which represents a fact about the managed entity or system.

158. What is data cleaning?
Ans: Data cleansing or **data scrubbing** is the act of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database.

159. What is clustering?
Ans: Grouping of similar objects.

160. What is Data purging?
Ans: The process of cleaning junk data is termed as data purging. Purging data would mean getting rid of unnecessary NULL values of columns

161. What are CUBES?
Ans: A data cube stores data in a summarized version which helps in a faster analysis of data. The data is stored in such a way that it allows reporting easily

162. What are OLAP and OLTP?
Ans: OLTP: Online Transaction and Processing helps and manages applications based on transactions involving high volume of data. Typical example of a transaction is commonly observed in Banks, Air tickets etc. Because OLTP uses client server architecture, it supports transactions to run cross a network.
OLAP: Online analytical processing performs analysis of business data and provides the ability to perform complex calculations on usually low volumes of data. OLAP helps the user gain an insight on the data coming from different sources (multi dimensional).

163. What is a SNAPSHOT?
Ans: Snapshots are read-only copies of a master table located on a remote node which is periodically refreshed to reflect changes made to the master table.

164. What are the different problems that "Data mining" can solve?
1. Data mining helps analysts in making faster business decisions which increases revenue with lower costs.
2. Data mining helps to understand, explore and identify patterns of data.
3. Data mining automates process of finding predictive information in large databases.
4. Helps to identify previously hidden patterns.

165. What are the different stages of data mining?
Ans
Exploration: This stage involves preparation and collection of data. it also involves data cleaning, transformation. Based on size of data, different tools to analyze the data may be required. This stage helps to determine different variables of the data to determine their behavior.
Model building and validation: This stage involves choosing the best model based on their predictive performance. The model is then applied on the different data sets and compared for best performance. This stage is also called as pattern identification. This stage is a little complex because it involves choosing the best pattern to allow easy predictions.
Deployment: Based on model selected in previous stage, it is applied to the data sets. This is to generate predictions or estimates of the expected outcome.

166. What is Discrete and Continuous data in Data mining world?
Ans: Discrete data
- Only finite set of values are available ex: zip codes, no. of words etc.
- Integer variables are represented very often.

Continuous data
- Only real numbers are available – ex : height, weight, length, temperature
- Represents floating point variables.

167. What is MODEL in Data mining world?
Ans: Models in Data mining help the different algorithms in decision making or pattern matching. The second stage of data mining involves considering various models and choosing the best one based on their predictive performance.

168. How does the data mining and data warehousing work together?
Ans: Data warehousing can be used for analyzing the business needs by storing data in a meaningful form. Using Data mining, one can forecast the business needs. Data warehouse can act as a source of this forecasting.

170. What is a Decision Tree Algorithm?
Ans: A decision tree is a tree in which every node is either a leaf node or a decision node. This tree takes an input an object and outputs some decision. All Paths from root node to the leaf node are reached by either using AND or OR or BOTH. The tree is constructed using the regularities of the data. The decision tree is not affected by Automatic Data Preparation.

171. What is Time Series algorithm in data mining?
Ans: Time series algorithm can be used to predict continuous values of data. Once the algorithm is skilled to predict a series of data, it can predict the outcome of other series. The algorithm generates a model that can predict trends based only on the original dataset. New data can also be added that automatically becomes a part of the trend analysis.
E.g. Performance one employee can influence or forecast the profit.

172. What is Sequence clustering algorithm?
Ans: Sequence clustering algorithm collects similar or related paths, sequences of data containing events. The data represents a series of events or transitions between states in a dataset like a series of web clicks. The algorithm will examine all probabilities of transitions and measure the differences, or distances, between all the possible sequences in the data set. This helps it to determine which sequence can be the best for input for clustering.

173. Explain how to mine an OLAP cube.
Ans: A data mining extension can be used to slice the data the source cube in the order as discovered by data mining. When a cube is mined the case table is a dimension.

174. What is an index?
Ans: Indexes help us to find data faster. It can be created on a single column or a combination of columns. A table index helps to arrange the values of one or more columns in a specific order.

Micro Processors

175. What is a Microprocessor?
Ans: Microprocessor is a program-controlled device, which fetches the instructions from memory, decodes and executes the instructions. Most Micro Processors are single- chip devices.

176. What are the flags in 8085?
Ans: Accumulator register, Temporary register, Instruction register, Stack Pointer, Program Counter are the various registers in 8085

182. **What is meant by Maskable interrupts?**
Ans: An interrupt that can be turned off by the programmer is known as Maskable interrupt.

183. What is Non-Maskable interrupts?
Ans: An interrupt which can be never be turned off (ie. disabled) is known as Non-Maskable interrupt

184. **What are the various segment registers in 8086?**
Ans: Code, Data, Stack, Extra Segment registers in 8086.

185. What are the flags in 8086?
Ans: In 8086 Carry flag, Parity flag, Auxiliary carry flag, Zero flag, Overflow flag, Trace flag, Interrupt flag, Direction flag, and Sign flag.
186. How many pins are there in 8086 micro processor?
Ans: 40 pins

187. What does EU do?
Ans: Execution Unit receives program instruction codes and data from BIU, executes these instructions and store the result in general registers.

188. **What are SIM and RIM instructions?**
Ans: SIM is Set Interrupt Mask. It is used to mask the hardware interrupts.
RIM is Read Interrupt Mask. Used to check whether the interrupt is Masked or not.

189. **What is the difference between 8086 and 8088?**
Ans: The BIU in 8088 is 8-bit data bus & 16- bit in 8086.Instruction queue is 4 byte long in 8088and 6 byte in 8086.

190. **Give example for Non-Maskable interrupts?**
Ans: Trap is known as Non-Maskable interrupts, which is used in emergency condition.

191. Give examples for 8 / 16 / 32 bit Microprocessor?
Ans: 8-bit Processor - 8085 / Z80 / 6800; 16-bit Processor - 8086 / 68000 / Z8000; 32-bit Processor - 80386 / 80486

192. What is meant by a bus?
Ans: A bus is a group of conducting lines that carriers data, address, & control signals.

193. Name 5 different addressing modes?
Ans: Immediate, Direct, Register, Register indirect, Implied addressing modes

194. What is the difference between primary & secondary storage device?
Ans: In primary storage device the storage capacity is limited. It has a volatile memory. In secondary storage device the storage capacity is larger. It is a nonvolatile memory. Primary devices are: RAM / ROM. Secondary devices are: Floppy disc / Hard disk.

195. What is Stack Pointer?
Ans: Stack pointer is a special purpose 16-bit register in the Microprocessor, which holds the address of the top of the stack.

196. In 8085 name the 16 bit registers?
Ans : Stack pointer and Program counter all have 16 bits.

197. What happens when HLT instruction is executed in processor?
Ans: The Micro Processor enters into Halt-State and the buses are tri-stated.

198. Which interrupt has the highest priority?
Ans: TRAP has the highest priority

199. What are Hardware interrupts?
Ans: TRAP, RST7.5, RST6.5, RST5.5, INTR

200. What is the difference between timer and counter?
Ans: A counter accumulates an unknown quantity of external events over a known interval of time. A timer accumulates a series events of a known interval over an interval that is being measured.