



Dr.G.R.Damodaran College of Science

(Autonomous, affiliated to the Bharathiar University, recognized by the UGC) Re-accredited at the 'A' Grade Level by the NAAC and ISO 9001:2008 Certified CRISL rated 'A' (TN) for MBA and MIB Programmes

III MBA (PART TIME) [2015-2018]
SEMESTER V

ELECTIVE : DATA WAREHOUSING AND DATA MINING - 552X1
Multiple Choice Questions.

1. Which of the following is true?

- A. The data warehouse consists of data marts and operational data
- B. The Data Warehouse consists of data marts and application data
- C. The Data Warehouse is used as a source for the operational data
- D. The operational data are used as a source for the data warehouse

ANSWER: B

2. How do you better define a data warehouse as

- A. Can be updated by end users.
- B. Contains numerous naming conventions and formats.
- C. Organized around important subject areas
- D. Contains only current data.

ANSWER: C

3. Which of the following is an operational system

- A. A system that is used to run the business in real time and is based on historical data
- B. A system that is used to run the business in real time and is based on current data
- C. A system that is used to support decision making and is based on current data
- D. A system that is used to support decision making and is based on historical data

ANSWER: B

4. The generic two-level data warehouse architecture includes

- A. at least one data mart.
- B. data that can be extracted from numerous internal and external sources
- C. near off-time updates.
- D. historic data.

ANSWER: B

5. Which of the following is reconciled data

- A. Current data intended to be the single source for all decision support systems
- B. Data stored in the various operational systems throughout the organization
- C. Data stored in one operational system in the organization
- D. Data that has been selected and formatted for end-user support applications

ANSWER: A

6. Which of the following is transient data

- A. Data in which changes to existing records cause the previous version of the records to be eliminated.
- B. Data in which changes to existing records do not cause the previous version of the records to be eliminated.
- C. Data that are never altered or deleted once they have been added

D. Data that are never deleted once they have been added
ANSWER: B

7. Which of the following is an extract process
- A. Capturing all of the data contained in various operational systems.
 - B. Capturing a subset of the data contained in various operational systems
 - C. Capturing all of the data contained in various decision support systems
 - D. Capturing a subset of the data contained in various decision support systems

ANSWER: B

8. What is Data Scrubbing
- A. A process to upgrade the quality of data before it is moved into a data warehouse
 - B. A process to reject data from the data warehouse and to create the necessary indexes
 - C. A process to upgrade the quality of data after it is moved into a data warehouse
 - D. A process to load the data in the data warehouse and to create the necessary indexes

ANSWER: A

9. The load and index is
- A. a process to upgrade the quality of data before it is moved into a data warehouse
 - B. a process to upgrade the quality of data after it is moved into a data warehouse
 - C. a process to reject data from the data warehouse and to create the necessary indexes
 - D. a process to load the data in the data warehouse and to create the necessary indexes

ANSWER: D

10. KDD describes the
- A. whole process of extraction of knowledge from data
 - B. extraction of data
 - C. extraction of information
 - D. extraction of rules

ANSWER: A

11. Translation of problem to learning technique is called as
- A. reengineering
 - B. translational engineering
 - C. representational engineering
 - D. learning algorithm

ANSWER: C

12. . Redundancy refers to the elements of a message that can be derived from other parts of
- A. different message
 - B. irrelevant message
 - C. same message
 - D. complete message

ANSWER: C

13. Metadata describes
- A. contents of database
 - B. structure of contents of database
 - C. structure of database
 - D. database itself

ANSWER: B

14. The partition of overall data warehouse is
- A. database.

- B. data cube
- C. data mart.
- D. operational data

ANSWER: C

15. _____ is used to load the information from operational database

- A. Replication technique
- B. Reengineering technique
- C. Engineering technique
- D. Transformation engineering

ANSWER: A

16. _____ multiprocessing machines share same hard disk and internal memory

- A. Massively parallel
- B. Symmetric
- C. Parallel
- D. Asymmetric

ANSWER: B

17. OLAP stands for

- A. Online Analytical Processing
- B. Online Linear Analytical Processing
- C. Online Animated Process
- D. Online Analytical Problem

ANSWER: A

18. OLAP is used to explore the _____ knowledge

- A. shallow
- B. deep
- C. multidimensional
- D. hidden

ANSWER: C

19. Hidden knowledge can be found by using

- A. searching algorithm
- B. pattern recognition algorithm
- C. searching algorithm
- D. clues

ANSWER: B

20. Deep knowledge can be found only by using

- A. clues
- B. OLAP
- C. SQL.
- D. any algorithm

ANSWER: A

21. The next stage to data selection in KDD process

- A. enrichment.
- B. coding.
- C. cleaning
- D. reporting

ANSWER: D

22. Enrichment means

- A. deleting data
- B. cleaning data
- C. adding external data
- D. selecting the data.

ANSWER: C

23. The decision support system is used only for

- A. cleaning
- B. coding
- C. selecting
- D. queries

ANSWER: D

24. In _____ approach data ware house is build first and all information needed is selected

- A. top-down
- B. . client/server
- C. bottom-up
- D. DSS

ANSWER: A

25. _____ is a creative activity that has to be performed repeatedly in order to get best results.

- A. Cleaning
- B. Reporting
- C. Coding
- D. Selection

ANSWER: C

26. The complexity of data mining algorithm is represented by

- A. $2n \log n$
- B. $3n \log n$
- C. $n \log n$
- D. $2 \log n$

ANSWER: C

27. _____ is the heart of knowledge discovery in database process

- A. Selection.
- B. Data ware house
- C. Data mining
- D. Creative coding

ANSWER: D

28. EIS stands for

- A. Exchange of Information System
- B. Extra Information System
- C. Executive Information System
- D. Extended Information system

ANSWER: C

29. . Foreign key constraints are also referred as

- A. consistency constraints
- B. referential integrity
- C. conditional integrity
- D. domain constraints

ANSWER: B

30. The set of attribute in a database that refers to data in another table is called _____
- A. primary key
 - B. candidate key
 - C. foreign key
 - D. constraints.

ANSWER: C

31. A database containing volatile data used for daily operation of an organization is _____
- A. historic data
 - B. metadata
 - C. knowledge
 - D. operational data

ANSWER: D

32. The system that can be used without knowledge of internal operation _____
- A. black box
 - B. white box
 - C. case based learning
 - D. Historical learning

ANSWER: A

33. . In KDD and data mining, noise is referred to as
- A. repeated data
 - B. complex data
 - C. meta data
 - D. random errors in database

ANSWER: D

34. DSS stands for
- A. Deciding Support System
 - B. Decision Support System
 - C. Decision Software System
 - D. Decision System of System

ANSWER: B

35. Data mining algorithms requires
- A. efficient sampling method
 - B. storage of intermediate results
 - C. capacity to handle large amounts of data
 - D. . All of the above

ANSWER: D

36. The technique of learning by generalizing from examples is
- A. incremental learning
 - B. inductive learning
 - C. hybrid learning
 - D. generalized learning

ANSWER: B

37. Knowledge discovery in database refers to
- A. whole process of extraction of knowledge from data
 - B. selection of data

- C. coding
- D. cleaning the data

ANSWER: A

38. Data mining is used to refer _____ stage in knowledge discovery in database

- A. selection
- B. retrieving
- C. discovery
- D. coding

ANSWER: C

39. Query tools and data mining tools are

- A. same
- B. different
- C. complementary
- D. standard

ANSWER: C

40. In genetic algorithm the problem is considered in terms of

- A. values
- B. points in multidimensional space
- C. node
- D. strings of characters

ANSWER: D

41. A _____ is necessary condition for KDD's effective implement

- A. data set
- B. database
- C. data warehouse
- D. data

ANSWER: C

42. The first international KDD conference was held in the year

- A. 1995
- B. 1994
- C. 1993
- D. 1992

ANSWER: A

43. AI stands for

- A. art of interest
- B. artificial interest
- C. art of intelligence
- D. artificial intelligence

ANSWER: D

44. KDD is a

- A. new technology that is use to store data
- B. multidisciplinary field of research
- C. database technology
- D. expert system

ANSWER: B

45. _____ could generate rule automatically

- A. KDD
- B. machine learning
- C. artificial intelligence
- D. expert system

ANSWER: B

46. Intelligent miner is a mining tool from

- A. Clementine.
- B. Microsoft
- C. IBM
- D. Wipro

ANSWER: C

47. _____ is a mining tool from integral solutions

- A. Thought
- B. web miner
- C. rapid miner
- D. clementine

ANSWER: D

48. A good introduction to machine learning is the idea of

- A. concept learning
- B. content learning
- C. theory of falsification
- D. Pople's law

ANSWER: A

49. . The algorithms that are controlled by human during their execution is _____ algorithm

- A. unsupervised
- B. batch learning
- C. supervised.
- D. incremental

ANSWER: C

50. Background knowledge depends on the form of

- A. theoretical knowledge
- B. hypothesis
- C. formulae
- D. knowledge representation

ANSWER: D

51. A _____ algorithm takes a new piece of information at each learning cycle and tries to revise the theory using new data

- A. supervised
- B. batch learning
- C. unsupervised
- D. incremental

ANSWER: D

52. In KDD process _____percentage is about mining

- A. 40
- B. 30
- C. 20
- D. 10

ANSWER: C

53. The easiest way to gain access to the data and facilitate effective decision making is to set up a _____.

- A. database
- B. data mart
- C. data warehouse
- D. operation al

ANSWER: C

54. Smaller local data warehouse is called as

- A. data mart
- B. database
- C. data model
- D. meta data

ANSWER: B

55. Data warehouse is only used for

- A. operating the data
- B. managing the data
- C. decision making
- D. queries

ANSWER: D

56. The _____ data are stored in data warehouse

- A. optimized
- B. maximized
- C. compressed
- D. historical

ANSWER: D

57. A decision support system is a system that

- A. can constantly change over time
- B. can not change
- C. copies the data
- D. supports the system

ANSWER: A

58. The _____ techniques are used to load information from operational database to data warehouse

- A. reengineering
- B. reverse.
- C. transfer
- D. replication

ANSWER: D

59. The _____ represents the best choice for building a data warehouse

- A. Client / server
- B. database
- C. bottom up
- D. visualization

ANSWER: A

60. . A _____ is not the rule that govern the basic structure of data warehouse.

- A. time dependent
- B. volatile
- C. subject oriented
- D. integrated

ANSWER: B

61. . Expert systems are

- A. system that contain the knowledge of specialists
- B. system that can think of their own
- C. system that can work
- D. system that can create the knowledge

ANSWER: A

62. A _____ is an interactive system that enables decision makers to use database and models on a computer in order to solve ill structured problems

- A. database
- B. client/server
- C. DSS
- D. mainframe

ANSWER: C

63. The _____ is a symbolic representation of facts or ideas from which information can potentially be extracted

- A. knowledge
- B. data
- C. algorithm
- D. program

ANSWER: B

64. A collection of interesting and useful patterns in database is called

- A. algorithm
- B. data
- C. information
- D. knowledge

ANSWER: D

65. In data mining software that works on local workstation is used to

- A. write coding
- B. generate screen and reports for the end user
- C. make decisions
- D. find patterns

ANSWER: B

66. The _____ operation is used for reducing data cube by one or more dimensions

- A. drilling
- B. rolling.
- C. dicing
- D. slicing

ANSWER: D

67. The main organizational justification for implementing a data warehouse is to provide _____

- A. cheaper ways of handling transportation
- B. decision support
- C. storing large volume of data

D. access to data
ANSWER: C

68. .KDD consists of _____ stages

- A. four
- B. five
- C. six
- D. seven

ANSWER: C

69. _____ is the first stage in KDD process

- A. Data selection
- B. Cleaning
- C. Mining
- D. Enrichment

ANSWER: A

70. The term that is not associated with data cleaning process is

- A. domain consistency
- B. de-duplication
- C. disambiguation
- D. egmentation

ANSWER: D

71. In _____ process of KDD additional information can be added to the existing data.

- A. enrichment
- B. coding
- C. selecting
- D. cleaning

ANSWER: A

72. _____ is a type of coding operation that occurs frequently in KDD context

- A. Filtering.
- B. Visualization
- C. Flattening
- D. Replication

ANSWER: C

73. Data mining methodology states that in optimal situation data mining is an

- A. standard process
- B. complete process
- C. creative process
- D. ongoing process

ANSWER: D

74. The ___ algorithm can be applied in cleaning data

- A. search
- B. pattern recognition
- C. learning
- D. clustering

ANSWER: B

75. Data Transformation includes

- A. a process to change data from a detailed level to a summary level

- B. a process to change data from a summary level to a detailed level
- C. joining data from one source into various sources of data
- D. separating data from one source into various sources of data

ANSWER: A

76. _____ is called a multifield transformation

- A. converts data from fields into field
- B. converts data from double fields into multiple fields
- C. converts data from one field to one field
- D. converts data from one field into multiple fields

ANSWER: D

77. _____ is the goal of data mining

- A. To confirm that data exists
- B. To analyze data for expected relationships
- C. To explain some observed event or condition
- D. To create a new data warehouse

ANSWER: C

78. Which of the following is not one of the three main components in a decision support system

- A. Model
- B. Communications
- C. Data
- D. Dialogue

ANSWER: B

79. Business Intelligence and data warehousing is used for

- A. Forecasting
- B. Data Mining.
- C. Analysis of large volumes of product sales data
- D. All the above

ANSWER: D

80. Which of the following uses a series of logically related two-dimensional tables or files to store information in the form of a database

- A. Database
- B. Database management system
- C. Data warehouse
- D. None of the above

ANSWER: D

81. Which of the following is a data manipulation tool

- A. File generators
- B. Query by example tool
- C. Structure question language
- D. Structure query language

ANSWER: B

82. Which of the following is true of three-tier data warehouses

- A. Once created, the data marts will keep on being updated from the data warehouse at periodic times.
- B. Once created, the data marts will directly receive their new data from the operational databases
- C. The data marts are different groups of tables in the data warehouse
- D. A data mart becomes a data warehouse when it reaches a critical size

ANSWER: A

83. Which of the given technology is not well-suited for data mining

- A. Expert system technology
- B. Data visualization
- C. Technology limited to specific data types such as numeric data types
- D. Parallel architecture

ANSWER: C

84. The most common source of change data in refreshing a data warehouse is _____.

- A. queryable change data
- B. cooperative change data
- C. logged change data
- D. snapshot change data.

ANSWER: A

85. _____ are responsible for running queries and reports against data warehouse tables

- A. hardware
- B. software
- C. end users
- D. middleware

ANSWER: C

86. Query tool is meant for

- A. data acquisition
- B. information delivery
- C. information exchange
- D. Communication

ANSWER: A

87. Which of the following function involves data cleaning, data standardizing and summarizing

- A. Storing data
- B. Transforming data
- C. Data acquisition
- D. Data Access

ANSWER: B

88. Data compression is to compress the given data by encoding in terms of

- A. bytes
- B. bits
- C. cluster
- D. roup

ANSWER: C

89. Which of the following feature is supported by the physical design of data warehouse

- A. Support backup and recovery
- B. Planning
- C. Analysis
- D. Coding

ANSWER: A

90. . Classification rules are extracted from

- A. root node
- B. decision tree

- C. siblings
 - D. branches
- ANSWER: B

91. Which one of the following is not true about OLAP
- A. They create no new knowledge
 - B. OLAP is powerful that data mining tool
 - C. They cannot search for new solution.
 - D. OLAP tool store their data in special multidimensional format

ANSWER: B

92. Association rules are always defined on
- A. binary attribute
 - B. single attribute
 - C. relational database
 - D. multidimensional attribute

ANSWER: A

93. The _____ is a knowledge that can be found by using pattern recognition algorithm.
- A. .hidden knowledge
 - B. deep.
 - C. shallow.
 - D. multidimensional

ANSWER: A

94. _____ refers to extracting knowledge from larger amount of data.
- A. Data abstraction
 - B. Data warehouse
 - C. Database
 - D. Data mining

ANSWER: D

95. _____ is a knowledge discovery process
- A. Data warehousing
 - B. Data mining.
 - C. Data cleaning
 - D. Data transformation

ANSWER: C

96. Data warehouse architecture is based on
- A. DBMS.
 - B. RDBMS
 - C. Sybase.
 - D. SQL Server

ANSWER: B

97. Source data from the warehouse comes from
- A. ODS
 - B. TDS
 - C. MDDB
 - D. ORDBMS

ANSWER: A

98. _____ is a data transformation process

- A. Comparison.
 - B. Projection.
 - C. Selection.
 - D. Filtering
- ANSWER: D

99. MDDDB stands for

- A. Multiple data doubling.
- B. Multidimensional databases
- C. Multiple double dimension
- D. Multi-dimension doubling

ANSWER: B

100. _____ is data about data

- A. Metadata.
- B. Microdata
- C. Minidata
- D. Multidata

ANSWER: A

101. Which one of the following feature is provided by an effective data visualization tool

- A. Capability to update data.
- B. Capability to delete or alter data
- C. Capability in append data
- D. Capability to compare data

ANSWER: D

102. _____ are some popular OLAP tools

- A. Metacube, Informix
- B. Oracle Express, Essbase
- C. HOLAP.
- D. MOLAP

ANSWER: A

103. Capability of data mining is to build _____ models

- A. retrospective.
- B. interrogative.
- C. predictive.
- D. predictive.

ANSWER: D

104. _____ is a process of determining the preference of customer's majority.

- A. Association
- B. Preferencing.
- C. Segmentation
- D. Classification

ANSWER: B

105. Which of the following is a process to humanize the mass of data

- A. Visualizing.
- B. Segmentation
- C. Discovery
- D. Correction

ANSWER: A

106. Strategic value of data mining is

- A. cost-sensitive
- B. work-sensitive
- C. time-sensitive
- D. technical-sensitive

ANSWER: C

107. Which of the following is a data store that is a subsidiary of a data warehouse of integrated data

- A. Data mart
- B. Data store
- C. Database
- D. Data house

ANSWER: A

108. Exceptional reporting in data warehousing is otherwise called as

- A. exception.
- B. alerts
- C. errors.
- D. bugs.

ANSWER: B

109. _____ is an expensive process in building an expert system

- A. Analysis
- B. Study
- C. Design
- D. Information collection

ANSWER: D

110. The full form of KDD is

- A. knowledge database
- B. knowledge data house
- C. knowledge discovery in database
- D. knowledge data definition

ANSWER: C

111. Removing duplicate records is a process called

- A. recovery
- B. data cleaning
- C. data cleansing
- D. data pruning

ANSWER: B

112. _____ helps to integrate, maintain and view the contents of the data warehousing system

- A. Business directory
- B. Information directory
- C. Data dictionary
- D. Database.

ANSWER: B

113. Discovery of cross-sales opportunities is called

- A. segmentation
- B. visualization

- C. correction
- D. association

ANSWER: D

114. Data marts that incorporate data mining tools to extract sets of data are called

- A. independent data mart
- B. dependent data marts
- C. intra-entry data mart
- D. inter-entry data mart

ANSWER: B

115. The power of self-learning system lies in

- A. cost
- B. speed.
- C. accuracy.
- D. implicit

ANSWER: C

116. _____ is the task of modern scientists in scientific research

- A. Theoretical
- B. Prediction
- C. Observation
- D. Analysis

ANSWER: B

117. _____ helps in effective decision making and easy access to data.

- A. Data mining
- B. Data warehousing
- C. Data mart
- D. Decision support system

ANSWER: B

118. Finding patterns in an observation is called

- A. Analysis
- B. Observation
- C. Theory
- D. Prediction

ANSWER: A

119. Databases used for performing daily transactions are called

- A. traditional database
- B. operational database
- C. data mart
- D. data warehouse

ANSWER: B

120. _____ is a system that keeps on changing constantly

- A. Closed system
- B. Arithmetic model
- C. Decision support system
- D. None

ANSWER: C

121. . Data warehouse demands high speed machines and variety of _____ processes

- A. optimization
- B. batch.
- C. production
- D. sequential

ANSWER: A

122. Data mining application is a _____ type of system

- A. artificial intelligence
- B. expert
- C. decision support system
- D. statistical system

ANSWER: C

123. Implementation of KDD demands

- A. deployment.
- B. integration
- C. process re-engineering
- D. process upgradation

ANSWER: C

124. _____ system contains the knowledge of the specialists

- A. Open system
- B. Closed system
- C. Expert system
- D. Decision making system

ANSWER: C

125. _____-is a process involving time and interaction with specialists.

- A. Knowledge gaining
- B. Analysis
- C. Design.
- D. Coding.

ANSWER: A

126. A more appropriate name for data mining could be

- A. internet Mining
- B. data Warehouse Mining
- C. knowledge Mining
- D. database Mining

ANSWER: C

127. Pick the odd one out

- A. SQL.
- B. Data Warehouse
- C. Data Mining
- D. OLAP

ANSWER: A

128. Which of the two are closely related to each other

- A. association rules & classification
- B. classification & prediction
- C. classification & clustering
- D. association rules & clustering

ANSWER: D

129. _____ is the phase of real discovery in the knowledge discovery process.

- A. Data mining
- B. Data mart.
- C. Database.
- D. Data warehouse

ANSWER: A

130. _____ is a very important step for building a reliable data mining system.

- A. Assessing
- B. Defining.
- C. Locating.
- D. Validating.

ANSWER: D

131. Repairing of missing or invalid data falls in the scope of

- A. data cleansing
- B. data mining
- C. data copying
- D. data deletion

ANSWER: A

132. _____ is the highest priority process in building a data mining application.

- A. Trial.
- B. Debugging
- C. Quality assurance
- D. Coding

ANSWER: C

133. SQL helps to find

- A. the interesting data
- B. hidden information
- C. intermediate data
- D. data under constraints that are already known

ANSWER: D

134. Translation of problem to learning technique is called as

- A. reengineering
- B. translational engineering
- C. representational engineering
- D. learning algorithm

ANSWER: C

135. Metadata describes

- A. contents of database
- B. structure of contents of database
- C. structure of database
- D. database itself

ANSWER: B

136. The partition of overall data warehouse is

- A. database
- B. data cube
- C. data mart

D. operational data
ANSWER: C

137. The information on two attributes is displayed in _____ in scatter diagram.
- A. visualization space
 - B. scatter space
 - C. cartesian space.
 - D. interactive space

ANSWER: C

138. The _____ plays an important role in artificial intelligence
- A. programming skill
 - B. scheduling
 - C. planning
 - D. learning capabilities

ANSWER: D

139. ETL means
- A. Extract, Transfer, Load
 - B. Extract, Transform, Load
 - C. Extract, Transport, Lead
 - D. Extract, Transshipment, Load

ANSWER: A

140. The _____ is called a multi field transformation.
- A. conversion of data from one field into multiple fields
 - B. conversion of data from fields into field
 - C. conversion of data from double fields into multiple fields
 - D. conversion of data from one field to one field

ANSWER: A

141. Which of the following table type belongs to snowflake schema
- A. Free
 - B. Dimension
 - C. Double
 - D. Replicator

ANSWER: B

142. Which of the following information systems are used in the daily running of the business
- A. Operational planning systems
 - B. Transaction processing systems
 - C. Process control systems
 - D. Office automation systems

ANSWER: A

143. Materials Requirements Planning (MRP) software is an example of an information systems application in which of the following areas
- A. Operations Management
 - B. Accounting
 - C. Marketing
 - D. Human Resource

ANSWER: A

144. What is the role of document image processing (DIP) systems

- A. Cash flow forecasting
- B. Inventory control
- C. Conversion into digital format
- D. Producing customer invoices

ANSWER: C

145. Which data administration subsystem periodically backs up information contained in a database

- A. Concurrency control facilities
- B. Reorganization facilities
- C. Backup and recovery facilities
- D. Security management facilities

ANSWER: C

146. Which of the following is true of three-tier data warehouses

- A. Once created, the data marts will keep on being updated from the data warehouse at periodic times.
- B. Once created, the data marts will directly receive their new data from the operational databases.
- C. The data marts are different groups of tables in the data warehouse
- D. A data mart becomes a data warehouse when it reaches a critical size

ANSWER: A

147. What is true about the multidimensional model

- A. It typically requires less disk storage
- B. It typically requires more disk storage
- C. Typical business queries requiring aggregate functions take more time
- D. Typical business queries requiring aggregate functions take more time

ANSWER: B

148. Technology area associated with CRM is

- A. specialization
- B. generalization
- C. personalization
- D. summarization

ANSWER: C

149. The terms equality and roll up are associated with

- A. OLAP.
- B. visualization.
- C. data mart.
- D. decision tree

ANSWER: C

150. Formulation of theory is called

- A. analysis.
- B. observation
- C. hypothesis
- D. prediction

ANSWER: C