Data Warehouse and Mining

1. _____ is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of
management decisions.
A. Data Mining.
B. Data Warehousing.
C. Web Mining.
D. Text Mining.
ANSWER: B

2. The data Warehouse is_____.
A. read only.
B. write only.
C. read write only.
D. none.
ANSWER: A

3. Expansion for DSS in DW is_____.
A. Decision Support system.
B. Decision Single System.
C. Data Storable System.
D. Data Support System.
ANSWER: A

4. The important aspect of the data warehouse environment is that data found within the data warehouse
is_____.
A. subject-oriented.
B. time-variant.
C. integrated.
D. All of the above.
ANSWER: D

5. The time horizon in Data warehouse is usually _____.
A. 1-2 years.
B. 3-4years.
C. 5-6 years.
D. 5-10 years.

ANSWER: D
6. The data is stored, retrieved & updated in _____.

A. OLAP.
B. OLTP.
C. SMTP.
D. FTP.
ANSWER: B

7. _____describes the data contained in the data warehouse.
A. Relational data.
B. Operational data.
C. Metadata.
D. Informational data.
ANSWER: C

8. _____predicts future trends & behaviors, allowing business managers to make proactive, knowledge-driven decisions.
A. Data warehouse.
B. Data mining.
C. Datamarts.
D. Metadata.
ANSWER: B

9. _____ is the heart of the warehouse.
A. Data mining database servers.
B. Data warehouse database servers.
C. Data mart database servers.
D. Relational data base servers.
ANSWER: B

10. _____ is the specialized data warehouse database.
A. Oracle.
B. DBZ.
C. Informix.
D. Redbrick.
ANSWER: D

11. _____defines the structure of the data held in operational databases and used by operational applications.
A. User-level metadata.
B. Data warehouse metadata.
C. Operational metadata.
D. Data mining metadata.
ANSWER: C

12. _____ is held in the catalog of the warehouse database system.
A. Application level metadata.
B. Algorithmic level metadata.
C. Departmental level metadata.
D. Core warehouse metadata.
ANSWER: B

13. _____maps the core warehouse metadata to business concepts, familiar and useful to end users.
A. Application level metadata.
B. User level metadata.
C. Enduser level metadata.
D. Core level metadata.
ANSWER: A

14. _____consists of formal definitions, such as a COBOL layout or a database schema.
A. Classical metadata.
B. Transformation metadata.
C. Historical metadata.
D. Structural metadata.
ANSWER: A

15. _____consists of information in the enterprise that is not in classical form.
A. Mushy metadata.
B. Differential metadata.
C. Data warehouse.
D. Data mining.
ANSWER: A

16. . _____databases are owned by particular departments or business groups.
A. Informational.
B. Operational.
C. Both informational and operational.
D. Flat.
ANSWER: B

17. The star schema is composed of _____ fact table.
A. one.
B. two.
C. three.
D. four.
ANSWER: A

18. The time horizon in operational environment is _____.
A. 30-60 days.
B. 60-90 days.
C. 90-120 days.
D. 120-150 days.
ANSWER: B

19. The key used in operational environment may not have an element of_____.
A. time.
B. cost.
C. frequency.
D. quality.
ANSWER: A

20. Data can be updated in _____environment.
A. data warehouse.
B. data mining.
C. operational.
D. informational.
ANSWER: C

21. Record cannot be updated in _____.
A. OLTP
B. files
C. RDBMS
D. data warehouse
ANSWER: D

22. The source of all data warehouse data is the_____.
A. operational environment.
B. informal environment.
C. formal environment.
D. technology environment.
ANSWER: A

23. Data warehouse contains_____data that is never found in the operational environment.
A. normalized.
B. informational.
C. summary.
D. denormalized.
ANSWER: C

24. The modern CASE tools belong to _____ category.
A. a. analysis.
B. b.Development
C. c.Coding
D. d.Delivery
ANSWER: A

25. Bill Inmon has estimated_____of the time required to build a data warehouse, is consumed in
the conversion process.
A. 10 percent.
B. 20 percent.
C. 40 percent
D. 80 percent.
ANSWER: D

26. Detail data in single fact table is otherwise known as_____.
A. monoatomic data.
B. diatomic data.
C. atomic data.
D. multiatomic data.
ANSWER: C

27. _____test is used in an online transactional processing environment.
A. MEGA.
B. MICRO.
C. MACRO.
D. ACID.
ANSWER: D

28. _____ is a good alternative to the star schema.
A. Star schema.
B. Snowflake schema.
C. Fact constellation.
D. Star-snowflake schema.
ANSWER: C

29. The biggest drawback of the level indicator in the classic star-schema is that it limits_____.
A. quantify.
B. qualify.
C. flexibility.

D. ability.
ANSWER: C

30. A data warehouse is _____.
A. updated by end users.
B. contains numerous naming conventions and formats
C. organized around important subject areas.
D. contains only current data.
ANSWER: C

31. An operational system is _____.
A. used to run the business in real time and is based on historical data.
B. used to run the business in real time and is based on current data.
C. used to support decision making and is based on current data.
D. used to support decision making and is based on historical data.
ANSWER: B

32. The generic two-level data warehouse architecture includes _____.
A. at least one data mart.
B. data that can extracted from numerous internal and external sources.
C. near real-time updates.
D. far real-time updates.
ANSWER: C

33. The active data warehouse architecture includes _____
A. at least one data mart.
B. data that can extracted from numerous internal and external sources.
C. near real-time updates.
D. all of the above.
ANSWER: D

34. Reconciled data is _____.
A. data stored in the various operational systems throughout the organization.
B. current data intended to be the single source for all decision support systems.
C. data stored in one operational system in the organization.
D. data that has been selected and formatted for end-user support applications.
ANSWER: B

35. Transient data is _____.
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: A

36. The extract process is _____.

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

37. Data scrubbing is _____.

A. a process to reject data from the data warehouse and to create the necessary indexes.

B. a process to load the data in 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 upgrade the quality of data before it is moved into a data warehouse

ANSWER: D

38. The load and index is _____.

A. a process to reject data from the data warehouse and to create the necessary indexes.

B. a process to load the data in 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 upgrade the quality of data before it is moved into a data warehouse.

ANSWER: B

39. 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

40. _____ is called a multifield transformation.

A. Converting data from one field into multiple fields.

B. Converting data from fields into field.

C. Converting data from double fields into multiple fields.

D. Converting data from one field to one field.

ANSWER: A

41. The type of relationship in star schema is _____.

A. many-to-many.

B. one-to-one.

C. one-to-many.
D. many-to-one.
ANSWER: C

42. Fact tables are _____.
A. completely demoralized.
B. partially demoralized.
C. completely normalized.
D. partially normalized.
ANSWER: C

43. _____ is the goal of data mining.
A. To explain some observed event or condition.
B. To confirm that data exists.
C. To analyze data for expected relationships.
D. To create a new data warehouse.
ANSWER: A

44. Business Intelligence and data warehousing is used for _____.
A. Forecasting.
B. Data Mining.
C. Analysis of large volumes of product sales data.
D. All of the above.
ANSWER: D

45. The data administration subsystem helps you perform all of the following, except_____.
A. backups and recovery.
B. query optimization.
C. security management.
D. create, change, and delete information.
ANSWER: D

46. 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

47. _____ are responsible for running queries and reports against data warehouse tables.
A. Hardware.
B. Software.

C. End users.
D. Middle ware.
ANSWER: C

48. Query tool is meant for _____.
A. data acquisition.
B. information delivery.
C. information exchange.
D. communication.
ANSWER: A

49. Classification rules are extracted from _____.
A. root node.
B. decision tree.
C. siblings.
D. branches.
ANSWER: B

50. Dimensionality reduction reduces the data set size by removing _____.
A. relevant attributes.
B. irrelevant attributes.
C. derived attributes.
D. composite attributes.
ANSWER: B

51. _____ is a method of incremental conceptual clustering.
A. CORBA.
B. OLAP.
C. COBWEB.
D. STING.
ANSWER: C

52. Effect of one attribute value on a given class is independent of values of other attribute is called _____.
A. value independence.
B. class conditional independence.
C. conditional independence.
D. unconditional independence.
ANSWER: A

53. 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

54. Multidimensional database is otherwise known as_____.
A. RDBMS
B. DBMS
C. EXTENDED RDBMS
D. EXTENDED DBMS
ANSWER: B

55. Data warehouse architecture is based on _____.
A. DBMS.
B. RDBMS.
C. Sybase.
D. SQL Server.
ANSWER: B

56. Source data from the warehouse comes from _____.
A. ODS.
B. TDS.
C. MDDB.
D. ORDBMS.
ANSWER: A

57. _____ is a data transformation process.
A. Comparison.
B. Projection.
C. Selection.
D. Filtering.
ANSWER: D

58. The technology area associated with CRM is _____.
A. specialization.
B. generalization.
C. personalization.
D. summarization.
ANSWER: C

59. SMP stands for _____.
A. Symmetric Multiprocessor.
B. Symmetric Multiprogramming.
C. Symmetric Metaprogramming.

D. Symmetric Microprogramming.
ANSWER: A

60. _____ are designed to overcome any limitations placed on the warehouse by the nature of the
relational data model.
A. Operational database.
B. Relational database.
C. Multidimensional database.
D. Data repository.
ANSWER: C

61. _____ are designed to overcome any limitations placed on the warehouse by the nature of the
relational data model.
A. Operational database.
B. Relational database.
C. Multidimensional database.
D. Data repository.
ANSWER: C

62. MDDB stands for _____.
A. multiple data doubling.
B. multidimensional databases.
C. multiple double dimension.
D. multi-dimension doubling.
ANSWER: B

63. _____ is data about data.
A. Metadata.
B. Microdata.
C. Minidata.
D. Multidata.
ANSWER: A

64. _____ is an important functional component of the metadata.
A. Digital directory.
B. Repository.
C. Information directory.

D. Data dictionary.
ANSWER: C

65. EIS stands for _____.
A. Extended interface system.
B. Executive interface system.
C. Executive information system.
D. Extendable information system.
ANSWER: C

66. _____ is data collected from natural systems.
A. MRI scan.
B. ODS data.
C. Statistical data.
D. Historical data.
ANSWER: A

67. _____ is an example of application development environments.
A. Visual Basic.
B. Oracle.
C. Sybase.
D. SQL Server.
ANSWER: A

68. The term that is not associated with data cleaning process is _____.
A. domain consistency.
B. deduplication.
C. disambiguation.
D. segmentation.
ANSWER: D

69. _____ are some popular OLAP tools.
A. Metacube, Informix.
B. Oracle Express, Essbase.
C. HOLAP.
D. MOLAP.
ANSWER: A

70. Capability of data mining is to build _____ models.
A. retrospective.
B. interrogative.
C. predictive.

D. imperative.
ANSWER: C

71. _____ is a process of determining the preference of customer's majority.
A. Association.
B. Preferencing.
C. Segmentation.
D. Classification.
ANSWER: B

72. Strategic value of data mining is _____.
A. cost-sensitive.
B. work-sensitive.
C. time-sensitive.
D. technical-sensitive.
ANSWER: C

73. _____ proposed the approach for data integration issues.
A. Ralph Campbell.
B. Ralph Kimball.
C. John Raphlin.
D. James Gosling.
ANSWER: B

74. The terms equality and roll up are associated with _____.
A. OLAP.
B. visualization.
C. data mart.
D. decision tree.
ANSWER: C

75. Exceptional reporting in data warehousing is otherwise called as _____.
A. exception.
B. alerts.
C. errors.
D. bugs.
ANSWER: B

76. _____ is a metadata repository.
A. Prism solution directory manager.
B. CORBA.
C. STUNT.

D. COBWEB.
ANSWER: A

77. _____ is an expensive process in building an expert system.
A. Analysis.
B. Study.
C. Design.
D. Information collection.
ANSWER: D

78. The full form of KDD is _____.
A. Knowledge database.
B. Knowledge discovery in database.
C. Knowledge data house.
D. Knowledge data definition.
ANSWER: B

79. The first International conference on KDD was held in the year _____.
A. 1996.
B. 1997.
C. 1995.
D. 1994.
ANSWER: C

80. Removing duplicate records is a process called _____.
A. recovery.
B. data cleaning.
C. data cleansing.
D. data pruning.
ANSWER: B

81. _____ contains information that gives users an easy-to-understand perspective of the information stored in the data warehouse.
A. Business metadata.
B. Technical metadata.
C. Operational metadata.
D. Financial metadata.
ANSWER: A

82. _____ 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

83. Discovery of cross-sales opportunities is called _____.
A. segmentation.
B. visualization.
C. correction.
D. association.
ANSWER: D

84. 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

85. _____ can generate programs itself, enabling it to carry out new tasks.
A. Automated system.
B. Decision making system.
C. Self-learning system.
D. Productivity system.
ANSWER: D

86. The power of self-learning system lies in _____.
A. cost.
B. speed.
C. accuracy.
D. simplicity.
ANSWER: C

87. Building the informational database is done with the help of _____.
A. transformation or propagation tools.
B. transformation tools only.
C. propagation tools only.
D. extraction tools.
ANSWER: A

88. How many components are there in a data warehouse?
A. two.
B. three.
C. four.

D. five.
ANSWER: D

89. Which of the following is not a component of a data warehouse?
A. Metadata.
B. Current detail data.
C. Lightly summarized data.
D. Component Key.
ANSWER: D

90. _____ is data that is distilled from the low level of detail found at the current detailed leve.
A. Highly summarized data.
B. Lightly summarized data.
C. Metadata.
D. Older detail data.
ANSWER: B

91. Highly summarized data is _____.
A. compact and easily accessible.
B. compact and expensive.
C. compact and hardly accessible.
D. compact.
ANSWER: A

92. A directory to help the DSS analyst locate the contents of the data warehouse is seen in _____.
A. Current detail data.
B. Lightly summarized data.
C. Metadata.
D. Older detail data.
ANSWER: C

93. Metadata contains atleast _____.
A. the structure of the data.
B. the algorithms used for summarization.
C. the mapping from the operational environment to the data warehouse.
D. all of the above.
ANSWER: D

94. Which of the following is not a old detail storage medium?
A. Phot Optical Storage.
B. RAID.
C. Microfinche.

D. Pen drive.
ANSWER: D

95. The data from the operational environment enter _____ of data warehouse.
A. Current detail data.
B. Older detail data.
C. Lightly summarized data.
D. Highly summarized data.
ANSWER: A

96. The data in current detail level resides till _____ event occurs.
A. purge.
B. summarization.
C. archieved.
D. all of the above.
ANSWER: D

97. The dimension tables describe the _____.
A. entities.
B. facts.
C. keys.
D. units of measures.
ANSWER: B

98. The granularity of the fact is the _____ of detail at which it is recorded.
A. transformation.
B. summarization.
C. level.
D. transformation and summarization.
ANSWER: C

99. Which of the following is not a primary grain in analytical modeling?
A. Transaction.
B. Periodic snapshot.
C. Accumulating snapshot.
D. All of the above.
ANSWER: B

100. Granularity is determined by _____.
A. number of parts to a key.
B. granularity of those parts.
C. both A and B.
D. none of the above.
ANSWER: C