



**ONLINE ACADEMY SYSTEM**  
**WITH IQBAL AND ANEESA**

**Online Academy System**

**CS-614 (DATA Warehousing)**

**Prepared By: Sir Iqbal and Mam Aneesa**

1. If one or more records in a relational table do not satisfy one or more integrity constraint, then the data:
  - a. Is syntactically dirty
  - b. Is semantically dirty page160**
  - c. Has Coverage anomaly
  - d. Has extraction issue
2. If actual data structure does not conform to documented formats then it is called:
  - a. Syntactically dirty data
  - b. Semantically dirty data page160**
  - c. Coverage anomaly
  - d. Extraction issue
3. Change Data Capture (CDC) can be challenging task because:

- a. Aggregates don't change in real time
- b. Transformation of extracted data is difficult
- c. Identifying the recently modified data may be difficult**

**page150**

- d. Source systems may not support extraction of changed aggregates
4. One feature of Change Data Capture (CDC) is that:
- a. It pre-calculates changed aggregates
  - b. It loads the transformed data in real time
  - c. It only processes the data that has been changed

**d. It can automate the transformation of extracted data**

**page150**

5. In context of data warehouse, normally it becomes difficult to extract data from different sources because these sources are normally:

**a. Heterogeneous page140**

- b. Homogeneous
- c. Centralized
- d. Base lined

6. Which Logical Data Extraction has significant performance impacts on the data warehouse server?

**a. Incremental Extraction page 133**

- b. Full Extraction
- c. Partial Extraction
- d. Offline Extraction

7. Suppose in system A, the possible values of "marital status" attribute were

"Male" & "Female", however in data warehouse, the values stored were "M"

for male and "F" for female. The above scenario is an example of:

**a. One-to-One scalar transformation ok**

- b. One-to-many element transformation
- c. Many-to-one element transformation
- d. Many-to-Many element transformation

8. In Extract, Load, Transformation (ELT) process, you don't need to purchase

extra devices to achieve parallelism because:

- a. ELT does not support parallelism

**b. You already have parallel data warehouse servers page 128**

- c. You have already performed transformation
- d. You have already performed Loading

9. The Kimball's iterative data warehouse development approach drew on decades of experience to develop the \_\_\_\_\_.

**a. Business Dimensional Lifecycle page 188**

- b. Data Warehouse Dimension
- c. Business Definition Lifecycle

d. OLAP Dimension

10. Which of the following is NOT one of the advantages of changed data capture

(CDC) technique.

a. Flat Files are not required.

**b. Limited Query interface is required for data extraction. Page -152**

c. No incremental online I/O required for log tape

d. Extraction of Change Data occurs immediately

11. The most common use of range partitioning is on \_\_\_\_\_

a. Color

**b. Date Page - 211**

c. PhoneNo

d. Name

12. Flat files are one of the prevalent structures used in \_\_\_\_\_ data extraction.

a. Online

**b. Offline Page - 134**

c. Incremental

d. Full 13. A relation is said to be in first normal form (1NF) if it does not contain \_\_\_\_\_.

a. Single value column

b. Multivalued column

c. Derived Column

d. Composite Column

14. In a fully normalized database, too many \_\_\_\_\_ are required.

a. Values

**b. Joins Page - 49**

c. Queries

d. Conditions

15. Identify data warehouse query from the following

a. List of students belong to Lahore City

b. Total number of students that have 3.5 CGPA

c. Number of students studying CS614 Course

d. Factors which can affect student performance

16. Which is not an issue of ROLAP in the following?

**a. Slandered hierarchy of dimensions Page-92**

b. Non slandered conventions

c. Maintenance

d. Aggregations pitfalls

17. OLTP implementations are fully \_\_\_\_\_.

**a. Normalized Page-30**

b. Demoralized

c. Predictive

d. Additive

18. For large record spaces and large number of records, the run time of the

clustering algorithms is \_\_\_\_\_

**a. Prohibitive Page - 164**

b. Static

c. Exponential

d. Numerical 19. Which is not a basic task of data transformation?

**a. Aggregation Page - 135**

b. Selection

c. Splitting / joining

d. Conversion

20. Record referring to the same entity are represented in different formats in the

different data sets are represented erroneously. Thus duplicate records will

appear in the merge database. This problem is known as \_\_\_\_\_.

**a. Merge/ purge problem page - 168**

b. Duplication problem

c. redundant duplication problem

d. redundant problem

21. Assume a company with a multi million row customer table i.e. n rows. Checking for referential integrity (RI) using a naive approach would take

**a.  $O(n)$  page 190**

b.  $O(1)$

c.  $O(\log n)$

d. None of the given

22. Development of data warehouse is hard because data source are usually \_\_\_\_\_.

a. Structured and homogeneous

b. Unstructured and Heterogeneous

c. Structured and heterogeneous

**d. Unstructured and homogeneous Page - 31**

23. In data transformation, \_\_\_\_\_ is the rearrangement and simplification of individual fields to make them more useful for the data warehouse environment.

a. Aggregation

b. Selection

c. Splitting / Joining

d. Conversion

**Note: Enrichment is correct option as per page 136 but it does not exist here.**

24. Ad-hoc access of data warehouse means: \_\_\_\_\_

- a. That have predefined database access pattern
- b. That does not have predefined database access pattern Page – 18**

c. That could be accessed by any user  
d. That could not be accessed by any user 25. Which is not a/an data quality validation technique?

**a. Consistency integrity Page – 189**

- b. Referential integrity
- c. Attribute domain
- d. Using data quality rules

26. Which one among the following is not an advantage of horizontal splitting.

- a. Enhance Security
- b. Organize Tables for different queries.

**c. Increase I/O overhead Page -55**

- d. Fast data retrieval

127 The min operator is conservative in that it assign to the dimension an aggregate value on higher than the value of its weakest data quality.

**a. Min Page - 92**

- b. Max
- c. Min and max
- d. None of give

27. In which class of aggregate average function can be placed ?

**a. Algebraic Page - 120**

- b. Distributive
- c. Associative
- d. Holistic

28. In the data warehouse environment the data is \_\_\_\_\_.

- a. Subject-oriented google
- b. Time-oriented
- c. Both subject and time oriented
- d. Neither time-oriented nor subject-oriented

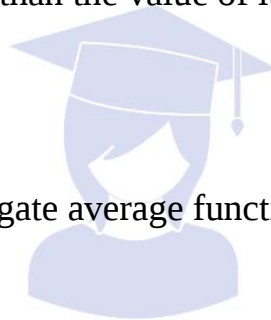
29. Online Extraction is a kind of \_\_\_\_\_ data extraction.

**a. Physical page 134**

- b. Logical
- c. Dimensional
- d. Multivalued

30. The \_\_\_\_\_ saw the advent of disk storage, or DASD (Direct Access Storage Device).

- a. 1960s
- b. 1970s page 13**
- c. 1950s



ONLINE ACADEMY SYSTEM

WITH IQBAL AND ANEESA

d. 1990s 31. Grain of a fact table means:

- a. The meaning of one fact table row page109**
- b. The meaning of one dimensional table row
- c. Summary of aggregates in all fact tables
- d. Summary of aggregates in all dimension tables

32. Normalization \_\_\_\_\_

- a. Reduces redundancy**
- b. Increases redundancy
- c. Reduces joins
- d. Reduces tables

33. Which of the following is NOT an example of a typical grain?

- a. Individual Transactions
- b. Daily aggregates
- c. Monthly aggregates

**d. Normalized attributes page111**

34. Multi-dimensional databases (MDDs) typically use \_\_\_\_\_ formats to store

pre-summarized cube structures.

- a. SQL
- b. Proprietary file page79**
- c. Object oriented
- d. Non-proprietary file

35. \_\_\_\_\_ provides a combination of "relational database access" and "cube" data

structures within a single framework.

**a. HOLAP page78**

- b. DOLAP
- c. MOLAP
- d. ROLAP

36. Data Warehouse provides the best support for analysis while OLAP carries out the \_\_\_\_\_ task.

- a. Mandatory
- b. Whole

**c. Analysis page69**

d. Prediction

37. \_\_\_\_\_ involves splitting a table by columns so that a group of columns is placed into the new table and the remaining columns are placed in another new table.

- a. Vertical splitting page56**
- b. Horizontal splitting
- c. Adding redundant column



ONLINE ACADEMY SYSTEM  
WITH IQBAL AND ANEESA

d. None of the given options 38.OLAP implementations are highly completely \_\_\_\_\_

a. Normalized

**b. Denormalized page69**

c. Predictive

d. Additive

39.If each cell of Relation contains a single value (no repeating values) then it is

confirmed that

**a. Relation R is in 1st Normal Form page 34**

b. Relation R is in 2nd Normal Form

c. Relation R is in 3rd Normal Form

d. Relation R is in 3rd Normal Form but not in 2nd Normal Form

40.Which kind of relationships is captured by Fact less fact table?

**a. Many-to-many page121**

b. One-to-many

c. One-to-one

d. None of the given options

41.Which of the following is NOT an example of dimension?

a. Product

b. Date

c. Region

**d. Sales Volume - VU Tech File**

42.Which people criticize Dimensional Modeling (DM) as being a data mart

oriented approach?

a. Those that consider ER model as Data marts

**b. Those that consider business processes as Data marts page110**

c. Those that consider Data mart as part of data warehouse

d. Those that consider dimensional modeling as denormalization approach

43.In a fully normalized form:

**a. Too many joins are required page49**

b. Relationships lose their significance

c. No joins are required

d. Data integrity becomes an issue

44.Which of the following is an example of Non-Additive Facts?

a. Quantity sold

b. Total Sale in Rs

**c. Discount in percentage page119**

d. Count of orders in a store 45.Which of the following is not a CUBE operation?

**a. ANSI SQL page80**

- b. Roll UP
- c. Drill Down
- d. pivoting

46. Allows download of "cube" structures to a desktop platform without the need

for shared relational or cube server.

- a. MOLAP
- b. ROLAP

**c. DOLAP page78**

- d. HOLAP

47. ROLAP provides access to information via a relational database using

**a. ANSI standard SQL page78**

- b. Proprietary file format
- c. Comma Separated Values
- d. All of the given options

48. \_\_\_\_\_ is usually deployed when expressions can be used to group data

together in such a way that access can be targeted to a small set of partitions for

a significant portion of the DW workload.

- a. Expression elimination
- b. Expression partitioning page67**
- c. Expression indexing
- d. None of the given option

49. Taken jointly, the extract programs or naturally evolving systems formed a

spider web, also known as:

- a. Distributed Systems Architecture
- b. Legacy Systems Architecture page14**
- c. Online Systems Architecture
- d. Intranet Systems Architecture

50. The data has to be checked, cleansed and transformed into a \_\_\_\_\_ format to

allow easy and fast access

**a. Unified page20**

- b. predicted
- c. qualified

d. proactive 51. Suppose in system A, the values of "Phone No" attribute were stored in

"country code-phone-extension" format, however after transformation in data

warehouse the separate columns were used for "country code", "phone" and

"extension". The above scenario is an example of:

- a. One-to-one scalar transformation
- b. One-to-many element transformation page 144**

- c. Many-to-one element transformation
- d. Many-to-Many element transformation

52. In decision support system ease of use is achieved by:

- a. Normalization
- b. Denormalization page 49**
- c. Drill up
- d. Drill down

53. Which of the following is one of the methods to simplify an ER model?

- a. Normalization
- b. Denormalization page 103**
- c. HOLAP
- d. Hybrid schema

54. In ETL process data transformation includes \_\_\_\_\_

- a. Data cleansing page 130**
- b. Data aggregation
- c. Behavior checking
- d. Pattern recognition

55. Non uniform use of abbreviations, units and values refers to:

- a. Syntactically dirty data page 160**
- b. Semantically dirty data
- c. Coverage anomaly
- d. Extraction issue

56. Suppose the size of the attribute "Computerized National Identity Card (CNIC)

no" is changed in NADRA database. This transformation refers to

- a. Format revision page 153**
- b. Field splitting
- c. Field decoding
- d. Calculation of derived value

57. The divide and conquer cube partitioning approach helps alleviate the \_\_\_\_\_ limitations of MOLAP implementation

- a. Flexibility
- b. Maintainability
- c. Security

**d. Scalability page 85** 58. Identify the TRUE statement(s) regarding Dimensional Modeling (DM):

- a. DM is inherently dimensional in nature
- b. DM comprises of a single central fact table
- c. DM comprises of a set of dimensional tables
- d. All of the given options page 103**

59. \_\_\_\_\_ can be used when some columns are rarely accessed rather than other columns or when the table has wide rows or header or both.

- a. Horizontal splitting
- b. Pre-joining

**c. Vertical splitting page56**

- d. Derived attributes

60. Which of the following is an example of derived attribute?

**a. Age page61**

- b. Size
- c. Color
- d. Length

61. The online high performance transaction processing was evolved in \_\_\_\_\_

- a. 1980

**b. 1975 page12**

- c. 1977
- d. 1965

62. Cube is a logical entity containing values of a certain fact at a certain aggregation level at an intersection of a combination of \_\_\_\_\_

- a. Facts

**b. Dimensions page88**

- c. Summary tables
- d. Primary and foreign keys

63. Which of the following statement is/are TRUE regarding Entity relationship modeling?

- a. It does not really model business; but models the micro relationships among data elements.

**b. ER modeling does not have "business rules," it has "data rules" page102**

- c. ER modeling helps retrieval of individual records having certain critical identifiers

- d. All of the given options 64. \_\_\_\_\_ facilitates a mobile computing paradigm.

**a. DOLAP page78**

- b. HOLAP
- c. ROLAP
- d. MOLAP

65. The main reason(s) for the increase in cube size may be

- a. Increase in the number of dimensions
- b. Increase in the cardinality of the dimensions

c. Increase in the amount of detail data

**d. All of the given options page87**

66. Suppose the amount of data recorded in an organization is doubled every year.

This increase is \_\_\_\_\_

a. Linear

b. Quadratic

**c. Exponential page15**

d. Logarithmic

67. The data in the data warehouse is \_\_\_\_\_

a. volatile

**b. Non-Volatile page20**

c. static

d. non-structured

68. \_\_\_\_\_ models the macro relationships among data elements with an

overall deterministic strategy

**a. Dimensional model page102**

b. Entity relationship model

c. Object oriented model

d. Structured model

69. De-normalization affects:

**a. Database size and query performance page52**

b. Database Usability and query reliability

c. Database availability and query success

d. None of the given options

70. \_\_\_\_\_ technique requires a separate column to specify the time and date

when the last modification was occurred.

a. Check Marks

**b. Time Stamps page150**

c. Just-In-Time

d. Real Time extraction

71. Which of the denormalization technique squeezes master table into detail?

**a. Pre-joining page58**

b. Horizontal splitting

c. Vertical splitting

d. Adding redundant column

72. De-normalization can help:

a. Minimize joins

b. Minimize foreign keys

c. Resolve aggregates

**d. All of the given options page51**

73.The domain of a "gender" field in some database may be {'F', 'M'} or as

{"Female", "Male"}) or even as {1, 0}. This is:

a. Primary key problem

**b. Non primary key problem page163**

c. Normalization problem

d. All of the given options

74.Increasing level of normalization \_\_\_\_\_ number of tables.

**a. Increases page51**

b. Decreases

c. Does not effect

d. None of the given options

75.Which of the following is not a Data Quality Validation Technique?

a. Referential Integrity

b. Using Data Quality Rules

c. Data Histogram

**d. Indexes page189**

76.This technique can be used when a column from one table is frequently

accessed in a large scale join in conjunction with a column from another table.

a. Horizontal splitting

b. Pre-joining

**c. Adding redundant column page58**

d. Derived attributes

77.Data cleansing requires involvement of domain expert because:

a. Domain expert has deep knowledge of data aggregation

b. Change Data capture requires involvement of domain expert

**c. Domain knowledge is required to correct anomalies page158**

d. Domain expert has deep knowledge of data summarization

78.Relational databases allow you to navigate the data in \_\_\_\_\_ that is

appropriate using the primary, foreign key structure within the data model.

a. Only One Direction

**b. Any Direction page19**

c. Two Direction

d. None of these

79.History is excellent predictor of the \_\_\_\_\_

a. Past

b. Present

**c. Future page15**

d. History

80. De-normalization is the process of selectively transforming normalized relations into un-normalized physical record specifications, with the aim to:

- a. Well structure the data
- b. Well model the data
- c. Reduce query processing time page 50**
- d. None of the given options

81. Which of the following is not a task of Data Transformation?

- a. Conversion
- b. Summarization
- c. Enrichment
- d. Full Data Refresh page 135, 154**

82. \_\_\_\_\_ gives total view of an organization.

- a. OLAP
- b. OLTP
- c. Data Warehouse page 15, 16**
- d. Database

83. Enrichment is one of the basic tasks in data \_\_\_\_\_

- a. Extraction
- b. Transformation page 138**
- c. Loading
- d. Summarization

84. Which of the following is not a technique of 'De-Normalization'?

- a. Pre-joining
- b. Splitting Tables
- c. Adding Redundant Columns
- d. ER Modeling page 50, 52**

85. Which of the following is an example of Additive Facts?

- a. Sales Amount page 119**
- b. Average
- c. Discount
- d. Ratios

86. Robotic libraries are needed for \_\_\_\_\_

- a. Cubes
- b. Data marts
- c. Data warehouse page 131**
- d. Aggregates

87. Normally ROLAP is implemented using \_\_\_\_\_

- a. Star Schema
- b. Hybrid Schema Page 78, 87**
- c. Pre-defined aggregates
- d. All of the given options

88.The relation R will be in 2nd Normal Form if

a. It is in 1NF and each cell contains single value.

**b. It is in 1NF and each non key attribute is dependent upon entire primary key. page44**

c. It is in 1NF and each non key attribute is dependent upon a single column

of composite primary key.

d. It is in 1NF and Primary key is composite.

89.In \_\_\_\_\_ nested-loop join of quadratic time complexity does not hurt the

performance

**a. Typical OLTP environments page22**

b. Data warehouse

c. DSS

d. OLAP

90.In Extract, Load, Transform (ELT) process, data transformation

**a. Takes place on the data warehouse server page147**

b. Takes place on a separate transformation server

c. Depends on the nature of the source database

d. Does not take place

91.Node of a B-Tree is stored in memory block and traversing a B-Tree involves

\_\_\_\_\_ page faults

a.o (n lg n)

**b. O ( lg n) page22**

c. O(n)

d. (n<sup>2</sup>)

92.As dimensions get less detailed (e.g. year vs. day) cubes get

**a. Smaller page84**

b. Larger

c. Partitioned

d. Merged

93.Which of the following is not a technique of "Changed Data Capture" in currently used Modern Source Systems?

a. Timestamps

b. Partitioning

c. Triggers

**d. Dimensional Modeling page150**

94.The trade-offs of denormalization is/are:

a. Storage

b. Performance

c. Ease-of-use

**d. All of the given options page62**

95.Data Warehouse is about taking / collecting data from different

\_\_\_\_\_ sources.

a. Harmonized

b. Identical

c. Homogeneous

**d. Heterogeneous page21**

96.It is observed that every year the amount of data recorded in an organization

**a. Doubles page15**

b. Triples

c. Quartiles

97.The users of data warehouse are knowledge workers in other words they are

\_\_\_\_\_ in the organization.

a. DWH Analyst

**b. Decision maker page18**

c. Database Administrator

d. Manager

98.In \_\_\_\_\_ system, the contents change with time.

**a. OLTP page20**

b. ATM

c. DSS

d. OLAP

99.The growth of master files and magnetic tapes exploded around the mid-

\_\_\_\_\_. **WITH IQBAL AND ANEESA**

**a. Mid-1960 page 12**

b. Mid-1970

c. Mid-1980

d. Mid-1990

100. Naturally Evolving architecture occurred when an organization had a

\_\_\_\_\_ approach to handling the whole process of hardware and software architecture.

**a. Relaxed page14**

b. Good

c. Not Relaxed

d. None

101. MDX by Microsoft is an example of \_\_\_\_\_

a. HOLAP

b. DOLAP

c. ROLAP

**d. None of the given options MOLAP page 79**

102. Which of the following is NOT an example of derived attribute?

a. Age

b. CGPA

c. Area of rectangle

**d. Height page60**

103. Table collapsing technique is applied in case of:

**a. One-to-one relation or many-to-many relation page52**

b. One-to-many relation

c. many to many relation

d. None of the given options

104. "Header size is reduced, allowing more rows per block, thus reducing I/O".

The above statement is TRUE with respect to:

**a. Vertical splitting page67**

b. Horizontal splitting

c. Adding redundant column

d. None of the given options

105. OLAP is:

a. Physical database design

b. Implementation technique

**c. Framework page69**

d. None of the given options 106. \_\_\_\_\_ Breaks a table into multiple tables based upon common column values

**a. Horizontal splitting page54**

b. Vertical Splitting

c. Adding redundant column

d. None of the given options

107. The issue(s) of "Adding redundant column" include(s):

a. Increase in table size

b. Maintenance

c. Loss of information

**d. All page60**

108. ROLAP provides access to information via a relational database using OLTP

implementations are fully\_\_\_\_\_.

**a. ANSI standard SQL page78**

b. Proprietary file format

c. Comma Separated Values

d. All

109. \_\_\_\_\_ is an application of information and data.

**a. Knowledge page11**

b. Intelligence

c. Power

d. Education

110. The need to synchronize data upon update is called

a. Data Imitation

b. Data Manipulation

c. Data Replication

**d. Data Coherency page12**

111. The input to the data warehouse can come from OLTP or transactional

system but not from other third party database.

a. True

**b. False page19**

112. A single database, couldn't serve both operational high performance transaction processing and DSS, analytical processing, all at the same time.

a. True

**b. False page13**

113. B-Tree is used as an index to provide access to records

**a. Without scanning the entire table page22**

b. By scanning the entire meta data

c. By scanning the entire table

d. None of these 114. For good decision making, data should be integrated across the organization

to cross the LoB (Line of Business). This is to give the total view of organization from:

a. Owner's Perspective

b. Decision Maker Perspective

**c. Customer's Perspective page16**

d. Employee's Perspective

115. Ad-hoc access means to run such queries which are known already.

a. True

**b. False page19**

116. Which statement is true for De-Normalization?

a. Redundant data is a performance liability at query time, but is a performance benefit at update time.

**b. Redundant data is a performance benefit at both query time and update time.**

c. Redundant data is a performance liability at both query time and update time.

d. Redundant data is a performance benefit at query time, but is a

performance liability at update time.

117. Pre-join technique is used to avoid

**a. Run time join page58**

b. Compile time join

c. Load time join

118. Cube is a \_\_\_\_\_ entity containing values of a certain fact at a certain

aggregation level at an intersection of a combination of dimensions.

**a. Logical page88**

b. Physical

c. Analytical

d. None of these

119. The goal of star schema design is to simplify \_\_\_\_\_

a. Logical data model

**b. Physical data model page107**

c. Conceptual data model

d. None of these

120. Grain is the \_\_\_\_\_ level of data stored in the warehouse.

**a. Atomic page111**

b. Summarized

c. Aggregated

d. Cube 121. Transactional fact tables do not have records for events that do not occur.

These are called

**a. Not Recording Facts page120**

b. Fact-less Facts

c. Null Facts

d. None of these

122. During ETL process of an organization, suppose you have data which can be

transformed using any of the transformation method. Which of the following

strategy will be your choice for least complexity?

**a. One-to-One Scalar Transformation**

b. One-to-Many Element Transformation

c. Many-to-Many Element Transformation

d. Many-to-One Element Transformation

123. Change Data Capture is one of the challenging technical issues in

**a. Data Extraction page149**

b. Data Loading

c. Data Transformation

d. Data Cleansing

124. Rearranging the grouping of source data, delivering it to the destination database, and ensuring the quality of data are crucial to the process of loading the data warehouse. Data \_\_\_\_\_ is vitally important to the overall health of a warehouse project.

- 1) Cleansing
- 2) Cleaning
- 3) Scrubbing

Which of the following options is true?

**a. Option 1 only page149**

- b. Option 2 only
- c. Option 1 & 2 only
- d. Option 1, 2 & 3

125. When performing objective assessments, companies follow a set of principles to develop metrics specific to their needs, there is hard to have “one

size fits all” approach. Which of the following statement represents the pervasive functional forms?

**a. Simple Ratio, Min or Max Operation, Weighted Average page186**

- b. Only Complex Ratio, Min Operation, Max Operation
- c. Only Simple Ratio, Min or Max Operation
- d. Only Min or Max Operation, Weighted Average

126. Normalization effects performance

- a. True
- b. False**

127. Collapsing tables can be done on the \_\_\_\_\_ relationships

- a. One-to-One
- b. Many-to-Many
- c. Both One-to-One and Many-to-Many page52**
- d. None of these

128. If  $w$  is the window size and  $n$  is the size of data set, then the complexity of

merging phase in BSN method is \_\_\_\_\_

- a.  $O(n)$
- b.  $O$
- c.  $O(w n)$
- d.  $O(w \log n)$  page171**

129. Relational modeling techniques are used to develop \_\_\_\_\_ systems.

- a. OLTP

**b. OLAP page 98**

c. MOLAP

d. ROLAP

130. Which of the following is an example of slowly changing dimensions?

a. Inheritance

b. Aggregation

c. Association

**d. Asset deposal pg no 124**

131. Time Complexity of Key Creation process in basic Sorted Neighborhood

(BSN) Method is \_\_\_\_\_

a.  $O(n \log n)$

b.  $O(\log n)$

**c.  $O(n)$  page 171**

d.  $O(2n)$

132. The degree to which values are present in the attributes that require them is

known as \_\_\_\_\_

**a. Completeness page 185**

b. Uniqueness

c. Accessibility

d. Consistency

133. \_\_\_\_\_ is the degree to which data accurately reflects the real-world object

that the data represents?

**a. Intrinsic Data Quality page 181**

b. Realistic Data Quality

c. Strong Data Quality

134. If a table is expected to have six columns but some or all of the records do

not have six columns then it is example of?

a. Syntactically dirty data page 160 but not found very clear

b. Semantically dirty data

c. Coverage anomaly

d. Extraction issue

135. One of the possible issues faced by web scrapping is that:

**a. Web pages may contain junk data page 146**

b. Web pages do not contain multiple facts

c. Web pages do not contain multiple dimensions

d. Web pages does not support transformation

136. An OLTP system is always good at \_\_\_\_\_

a. Evolving data page 122

- b. Keeping Static data
- c. Tracking past data
- d. Marinating historic data

137. \_\_\_\_\_ in database or data warehouse has no actual value, it only has potential value.

**a. Data page 181**

- b. Entity
- c. Flat table
- d. Data marts

138. The degree of similarity measured \_\_\_\_\_, different attributes may contribute differently.

**a. Numerically page 169**

- b. Qualitative
- c. Quantitative
- d. None of the given

139. \_\_\_\_\_ segregates data into separate partitions so that queries do not need to examine all data in the table.

- a. Pre-joining technique
- b. Collapsing tables technique
- c. Horizontal splitting technique page 55**
- d. Vertical splitting technique

140. The relation R will be in 3rd Normal Form if

- a. It is in 2NF and each cell contains single value
- b. It is in 2NF and every non-key column is non-transitively dependent upon its primary key. Page 46
- c. It is in 1NF and each non key attribute is dependent upon a single column of composite primary key.
- d. It is in 2NF and each non key attribute is dependent upon other non-key attribute.

141. OLAP is

**a. Analytical Processing page 69**

- b. Transaction Processing
- c. Additive processing
- d. Active processing

142. De-normalization usually speeds up \_\_\_\_\_.

- a. Data insertion

**b. Data retrieval page 51**

- c. Data deletion
- d. Data sharing

143. The hardware (CPU) utilization in data warehouse environment is full or

\_\_\_\_\_.

- a. Fixed
- b. Partial
- c. Not at all page 24**
- d. Slow

144. \_\_\_\_\_ is applicable in Profitability analysis.

- a. OLTP
- b. Data warehouse
- c. Information System (IS)
- d. Management Information System (MIS)

145. Time variant is a characteristic of data warehouse which means:

**a. Data loaded in data warehouse will be time stamped**  
**page 20**

- b. Data can be loaded in data warehouse anytime
- c. Data cannot be loaded in data warehouse with respect to time

146. In OLAP, the typical write operation is \_\_\_\_\_

**a. Bulk insertion page 75**

- b. Single insertion
  - c. Sequential insertion
147. In the decision support environment, the decision maker is interested in

\_\_\_\_\_.

- a. Only limited organizational data
- b. Big picture of organizational data page 21
- c. Only sale related data
- d. Only customer related data

148. \_\_\_\_\_ can be created in operational systems to keep track of recently updated records.

- a. Triggers page 150
- b. Timestamps
- c. Partitioning
- d. ETL

149. \_\_\_\_\_ is not the characteristic of data warehouse.

- a. Time variant
- b. Subject-oriented
- c. Integrated
- d. Volatile google

150. Consider the following Employee table and identify the column which

causes that the table is not in first normal forms (1NF).

- a. Emp\_ID

- b. Emp\_Name
- c. Emp\_skills
- d. Emp\_Designation

151. The application of data and information leads to \_\_\_\_\_.

- a. Intelligence
- b. Experience
- c. Knowledge
- d. Power

152. Which of the following is not an “Orr’s Law of Data quality”?

- a. “Data that is not used cannot be correct!”
- b. “Data quality is a function of its use, not its collection!”
- c. “Data will be no better than its most stringent use!”
- d. “Data duplication can be harmful for the organization!” page 181

153. Which type of dependency is represented by the following functional dependencies?

Book\_Name  $\square$  Author\_Name, Author\_Age

Author\_Name  $\square$  Author\_Age

- a. Partial dependency
- b. Transitive dependency page 173 form CS403
- c. Full-functional dependency
- d. Multivalued dependency

154. Telecommunication data warehouse is dominated by the \_\_\_\_\_ volume

of data generated at the call level.

- a. Partial
- b. Incomplete
- c. Sheer page 135**
- d. Semi-complete

155. Simple scalar transformation is a \_\_\_\_\_ mapping from one set of values to

another set of values using straightforward rules.

**a. One-to-one page 144**

- b. One-to-many
- c. Many-to-many
- d. Many-to-one

156. Information can answers questions like “what”, “who”, and “when” while

knowledge can answer questions like\_\_\_\_\_.

- a. Why
- b. Where
- c. Which

**d. How page 11**

157. Normalization is used to reduce\_\_\_\_\_.

**a. Reduce redundancy page 41**

- a. Increase redundancy
- b. Reduces joins
- c. Reduces tables

158. In context of Change Data Capture (CDC), some time a \_\_\_\_ object can be used to store the recently modified data.

- a. Buffer table

**b. Change table page 149**

- c. Checkmark table
- d. Change control table

159. \_\_\_\_\_ should not be present in a relation, so that it would be in second normal form (2NF).

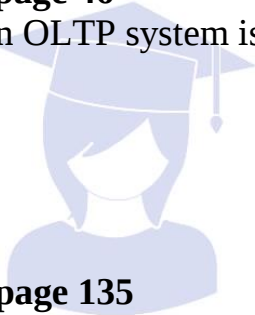
- a. Partial dependency
- b. Full-function dependency
- c. Multivalued dependency

**d. Transitive dependency page 46**

160. The data perspective in OLTP system is operational, while tat in data warehouse is \_\_\_\_\_.

- a. Fully Normalized
- b. Fully de-normalized
- c. Fully summarized

**d. Historical and detailed page 135**

161. The typical availability of OLTP system is 24/7, while that of data warehouse is \_\_\_\_\_.  


**a. 6/12 page 30**

- b. 7/12
- c. 7/24
- d. Twice a week

162. One goal of horizontal splitting is spreading rows of a table for exploiting \_\_\_\_\_.

- a. Sequentialism

**b. Parallelism page 54**

- c. Randomness
- d. All of the given options

163. Which is the most complex type of transformation in the following?

**a. Many-to-many element transformation page 144**

- b. One-to-one scalar transformation
- c. One-to-many element transformation
- d. All the given option

164. The last step of Software Development Life Cycle (SDLC) is implementation, while that of data warehouse is:

a. Integration

**b. Understanding requirements page 29**

c. Analysis

d. Testing

165. The \_\_\_\_\_ technique is a discipline used to highlight the microscopic relationships among data elements or entities.

**a. ER Modeling page 99**

b. DM Modeling

c. Relational Modeling

d. Multi-dimensional Modeling 166. Time Complexity of Basic Sorted Neighborhood (BSN) Method is

\_\_\_\_\_.

a.  $O(n \log n)$  page 171

b.  $O(\log n)$

c.  $O(n)$

d.  $O(2n)$

167. In OLTP environments, the size of tables is relatively \_\_\_\_\_.

a. Large

b. Fixed

c. Moderate

**d. Small page 22**

168. Depending on the data requirements of the warehouse, both summarization

and aggregation can be deployed during \_\_\_\_\_.

**a. Data transformation page 155**

b. Data Cleansing

c. Data extraction

d. Data loading

169. \_\_\_\_\_ is a logical design technique that seeks to present the data in a

standard that supports high-performance and ease of understanding.

**a. DM Page -103**

b. ER

c. Normalization

d. Relational Modeling

170. A/an \_\_\_\_\_ is a collection of random transactional codes flags and / or

text attributes that are unrelated to any particular dimension.

**a. Junk dimension Page 127**

b. Slowly Changing Dimension

c. Multi-valued Dimension

d. Simple Dimensions

171. \_\_\_\_\_ is the application of intelligence and experience to get common goals

**a. Wisdom Page - 11**

b. Education

c. Power d. Information

172. \_\_\_\_\_ is making all efforts to increase effectiveness and efficiency in meeting expected customer expectations.

a. Quality assurance

**b. Quality improvement Page -183**

c. Quality Maintenance

d. Quality Establishment

173. De-normalization is more like a “controlled crash” with the aim to \_\_\_\_\_ without loss of information.

a. Check

b. Balance

c. Decrease

**d. Enhance page - 49**

174. In the data warehouse data is collected from \_\_\_\_\_ sources.

a. Homogenous

**b. Heterogeneous Page -130**

c. External

d. Internal

175. The main goal of normalization is to eliminate \_\_\_\_\_.

**a. Data redundancy Page ref not found search from google**

b. Data sharing

c. Data Security

d. Data Consistency

176. One of the main reasons for the failure of DWH deployments is \_\_\_\_\_

**a. Data quality Page -179**

b. Data integrity

c. Data duplication

d. Data anomaly

177. In case of multiple sources for the same data element, we need to prioritize

the source systems on a per data element basis.

**a. Ranking Page - 143**

b. Prioritization

c. Element Selection

d. Measurement Selection 178. In OLTP environment the selectivity is \_\_\_\_\_ and \_\_\_\_\_ in data warehouse.

**a. High, Low page 22**

b. Low, High

c. High, Fixed

d. Fixed, Low

179. Data warehouse stores \_\_\_\_\_.

a. Operational data

**b. Historical Data Page -16**

c. Meta data

d. Log files data

180. OLAP implementations are highly / completely \_\_\_\_\_.

a. Normalized

**b. Denormalized Page 69**

c. Predictive

d. Additive

181. The business processes covered by ER diagram:

**a. Do not co-exist in time and space Page 109.**

b. Co-exist in time and space

c. Do not physically exist in real time context

d. None of the given options

182. For good decision making one should be able to integrate data across the

organization so as to cross the LoB (Line of Business).

a. Partial view

**b. Total view Page 16**

c. Undermine view

d. Brief view

183. Simple scalar transformation is a \_\_\_\_\_ mapping from one set of value to another.

**a. One-to-one Page 144**

b. One-to-many

c. Many-to-many

d. Many-to-one

184. Which of the following is/are example(s) of dimension?

a. Product

b. Region

c. Date

**d. All of the given options** 185. Which is not a class of anomalies in the following?

a. Dirty anomalies

**b. Syntactical Dirty date page 159**

c. Semantically Dirty data

d. Coverage anomalies

186. The biggest problem with \_\_\_\_\_ is the requirement of large main memory

as the cube size increase.

**a. MOLAP**

b. ROLAP

c. DATA Ware House

d. OLTP

187. Which one is the characteristic of data warehouse quires?

a. Use primary keys

b. High selectivity

**c. Use multiple tables page 30**

d. Return very rows

188. Which of the following is not an example of dimension?

a. Sales amount

**b. ATM Card no**

c. Time / Date

d. ATM \_ Location

189. ROLAP tools will query the relational data base using SQL generated to conform to framework using the fact and dimensions paradigm using the \_\_\_\_

**a. Star Schema page 87**

b. Snowflake

c. Relational schema

d. OPAL

190. One scope of data warehouse is to \_\_\_\_\_

**a. Improve business Page 30**

b. Run business

c. Record day-to-day business activities

d. Calculate tax of the business's profit 191. Which is not a/an characteristics of data quality?

**a. Reliability Page 185**

b. Uniqueness

c. Accessibility

d. Consistency

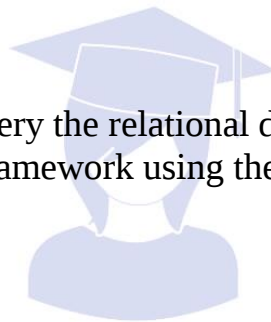
192. Which is not a/an distance Function in Basic Sorted Neighborhood (BSN)

Method?

**a. Matrix distance Page 176**

b. Phonetic distance

c. Typewriter distance



ONLINE ACADEMY SYSTEM

WITH TABAL AND ANEESA

d. Edit Distance

193. \_\_\_\_\_ should not be present in a relation, so that it would be in third

normal form (3NF).

**a. Transitive dependency Page 46**

b. Partial dependency

c. Full-functional dependency

d. Multivalued dependency

194. In which class of aggregates Max function can be placed?

**a. Distributive Page 120**

b. Holistic

c. Algebraic

d. Associative

195. There are often multiple ways to represent the same piece of information.

"FAST", "National University", "FAST NU" and "Nat. Univ. of Computers".

This problem is known as \_\_\_\_\_.

**a. Non primary key problem Page 163**

b. Primary key problem

c. Simple key problem

d. Composite key problem

196. The extent to which data is in appropriate languages, symbols, and units, and

the definitions are clear is known as \_\_\_\_\_.

**a. Interpretability Page 185**

b. Uniqueness

c. Accessibility

d. Consistency

197. In which type of data extraction, there is no need keep track of changes?

a. Full Extraction Page 133

b. Incremental Extraction

c. Full Extraction

d. Half Extraction **Note: Option A and Option C, both are the same in quiz.**

198. If a product meets formally defined "requirement specifications" yet fails to

be a quality product from the customer's perspective, this means the requirements were \_\_\_\_\_

**a. Defective page 180**

b. Unclear

c. Unrefined

d. Undefined

199. Which is not a/an step of data cleansing procedure?

a. Aggregation

b. Elementizing

c. Standardizing

d. Verifying

200. The end users of a data warehouse are \_\_\_\_\_

a. Programmers

b. Database developers

c. Data entry operators

d. Business executives

201. The hybrid OLAP (HOLAP) solution is a mix of architectures that supports queries against summary and transaction data in an integrated?

a. MOLAP and ROLAP

b. MOLAP

c. ROLAP

d. OLTP

202. Dirty data means that

a. Data cannot be aggregated

b. Data contains non-additive facts

c. Data does not fulfill dimensional modeling rules

d. Data does not conform to proper domain definitions page 158

203. \_\_\_\_\_ can result in costly errors, such as: False frequency distributions

and incorrect aggregates due to double counting.

a. Data duplication page 165

b. Data reduction

c. Data anomaly

d. Data transformation 204. A \_\_\_\_\_ is defined by a group of records that have similar

characteristics (behavior) for p% of the fields in the data set, where p is a user

defined value (usually above 90).

a. Pattern page 164

b. Cluster

c. Entity

d. Attribute

205. The \_\_\_\_\_ task is typically performed after most other transformation and cleaning steps have taken place.

a. Data Duplicate elimination page 165

b. Data transformation

c. Data extraction

d. Data loading

206. The \_\_\_\_\_ dimension represents data correctness

**a. Free-of-error page 187**

b. Completeness

c. Consistency

d. Correctness

207. One of the fundamental purpose of de normalization is to \_\_\_\_\_ a

number of physical tables which ultimately reduce the number joins to answer a

query.

a. Delete

b. Share

c. Increase

d. Decrease page 50

208. The process of converting entity relationship model in to dimensional model

comprises of \_\_\_\_\_ steps:

a. Two page 103

b. Three

c. Four page 109

d. Five

209. \_\_\_\_\_ does not have "business rules," it has "data rules."

**a. ER modeling page 102**

b. DM modeling

c. Multi-dimensional Modeling

d. OLAP

210. Instance matching between different sources is then achieved by a standard

\_\_\_\_\_ on the identifying attribute(s), if you are very, very lucky

**a. Equi-join page 169**

b. Inner join

c. Outer join

d. Fuller join

211.

1) Online Extraction is a kind of ----- data extraction:

- Physical page 132

- Logical

- Dimensional

- Multi valued

2) The ----- saw the advent of disk storage, or DASD( direct Access Storage Device) :

- 1960s

- 1970s page 13
- 1950s
- 1990s

3) In context of data warehouse, normally it becomes difficult to extract data from different

sources because these sources are normally.

- Heterogeneous page 140
- Homogeneous
- Centralized
- Baseline

4) Which of the following is not a task of Data Transformation?

- Conversion
- Summarization
- Enrichment

5) Which of the following is not an Orr's Law of Data Quality?"

- "Data that is not used cannot be corrected!"
- "Data quality is a function of its use, not its collection!"
- "Data will be no better than its most stringent use!"
- "Data duplication can be harmful for the organization! " page 181

6) Flat files are one of the prevalent structures used in ----- data extraction:

- Online
- Offline page 134
- Incremental

7) Which of the following is NOT one of the advantages of changed data capture (CDC) technique?

- Flat files are not required
- Limited query interface is required for data extraction page 152
- No incremental on-line I/O required for log tape
- Extraction of changed data occurs immediately

8) The most common range partitioning is on

- Color
- Date page 66
- PhoneNo
- Name

9) A relation is said to be in first normal form(1NF), if it does not contain

- Single value column
- Multi-valued column page 43
- Derived column

10) In a fully normalized database, too many \_\_\_\_\_ are required

- Values
- Joins page 49
- Queries
- Conditions

11) In the data warehouse, data is collection from ----- sources:

- Homogeneous
- Heterogeneous page 21
- External
- Internal

12) De-normalization is more like a “controlled crash” with the aim to --- without loss of information:

- Check
- Balance
- Decrease
- Enhance page 49

13) ----- is making all efforts to increase effectiveness and efficiency in meeting and accepted customer expectation:

- Quality assurance
- Quality improvement page 183
- Quality maintenance
- Quality Establishment

14) ----- is the application of intelligence and experience to get common goals. • Wisdom page 11

- Education
- Power
- Information

15) In the data transformation, ----- is the rearrangement and simplification of individual

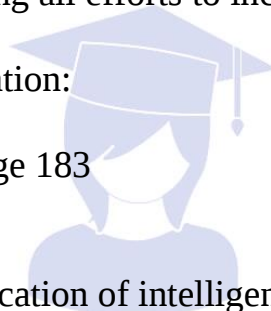
- Aggregation
- Enrichment page 136
- Splitting joining
- Conversion

16) Grain of a fact table means :

- The meaning of one fact table row page 109
- The meaning of one dimensional table row
- Summary of aggregates in all fact tables
- Summary of aggregates in all dimension tables

17) Normalization ----- :

- Reduces redundancy page 41
- Increases redundancy



ONLINE ACADEMY SYSTEM  
WITH IQBAL AND ANEESA

- Reduces joins
- Reduces tables

18) Which of the following is not an example of a typical grain :

- Individual transaction
- Daily aggregates
- Monthly aggregates
- Normalized attributes page 111•

19) Multi-dimensional databases(MDDs) typically use ----- formats to store pre summarized cube structures:

- SQL
- Proprietary file page 79
- Object oriented
- Non-proprietary file

20) ----- provides a combination of “relational databases access” and “cube” data structures within a single framework:

- HOLAP page 78
- DOLAP
- MOLAP
- ROLAP

21) Data Warehouse provides the best support for analysis while OLAP carries out the ----- task:

- Mandatory
- Whole
- Analysis page 69
- Prediction

22) ----- involves splitting a table by columns so that a group of columns is placed into the new table and the remaining columns are placed in another new table:

- Vertical splitting page 56
- Horizontal splitting
- Adding redundant column
- None of the given option

23) OLAP implementations are highly/completely ----- :

- Normalized
- Demoralized page 69• Predictive
- Additive

24) If each cell of Relation R contains a single value ( no repeating values) then it is confirmed that :

- Relation R is in 1st Normal Form page 43
- Relation R is in 2nd Normal Form



ONLINE ACADEMY SYSTEM  
WITH IQBAL AND ANEESA

- Relation R is in 3rd Normal Form
  - Relation R is in 3rd Normal Form but not in 2nd Normal Form
- 25) Which kind of relationships is captured by Fact less fact table:
- Many- to- Many page 121
  - One-to-many
  - One-to-one
  - None of the given option
- 26) Which of the following is not an example of dimension:
- Product
  - Date
  - Region
  - Sales volume page 78
- 27) Which people criticize Dimensional Modeling (DM) as being a data mart oriented approach?
- Those that consider ER models as Data marts
  - Those that consider Business processes as Data marts page 110
  - Those that consider Data marts as Data warehouse
  - Those that consider dimensional model
  - Those that consider dimensional modeling as de-normalization approach
- 28) In a fully normalized form:
- To many joins are required page 49
  - Relationships lose their significance
  - No joins are required
  - Data integrity becomes an issue
- 29) Which of the following is an example of Non-Additive Facts:
- Quality sold
  - Total sale in Rs.
  - Discount in percentage page 119
  - Count of orders in a store
- 30) Which of the following is not a CUBE operation?
- ANSI SQL page 81
  - Roll Up
  - Drill Down
  - Pivoting
- 31) ----- allows download of “cube” structures to a desktop platform without the need for shared or cube server:
- MPLAP
  - ROLAP
  - DOLAP page 78
  - HOLAP
- 32) ROLAP provides access to information via a relational database using:
- ANSI standard SQL page 78

- Proprietary file format
- Comma Separated Values
- All of the given option

33) ----- is usually deployed when expression can be used to group data together in such

a way that access can be targeted to a small set of partitions:

- Expression elimination
- Expression partitioning page 67
- Expression indexing
- None of the given option

34) Taken jointly, the extract programs or naturally evolving systems formed a spider web, also

known as

- Distributed Systems Architecture
- Legacy System Architecture page 14
- Online System Architecture
- Intranet System Architecture

35) The data has to be checked , cleaned and transformed into a -----

---- format to allow easy

and fast access

- Unified page 20
- Predicated
- Qualified
- Proactive

36) Suppose in a system A, the values of “PhoneNo” attribute were stored in “countrycode-phone

extension” format, however after transformation into data warehouse the separate columns were used for “countrycode”, “phone” and “extension”.

The above scenario is an example of :

- One-to-one scalar transformation
- One-to-many element transformation page 144+conceptual

- Many-to-one element transformation
- Many-to-many element transformation

37) In decision support system ease of use is achieved by:

- Normalization
- Denormalization page no 49
- Drill up
- Drill down

38) Which of the following is one of the methods to simplify an ER model?

- Normalization
- Denormalization page no 103
- HOLAP
- Hybrid schema

39) In ETL process data transformation includes -----

- Data cleansing page 129
- Data aggregation
- Behavior checking
- Pattern recognition

40) Non-uniform use of abbreviations, units, and values refers to:

- Syntactically dirty data page 160
- Semantically dirty data
- Coverage anomaly
- Extraction issue

41) Suppose the size of the attribute “Computerized National Card (CNIC) no. is changed in NADRA database. This transformation refers to:

- Format revision page 153
- Field splitting
- Field decoding
- Calculation of derived value

42) The divide and conquer cube partitioning approach helps alleviate the ----- limitations of

MOLAP implementation: • Flexibility

- Maintainability
- Security
- Scalability page 85

43) identify the TRUE statement:

- DM is inherently dimensional in nature
- DM comprises of a single central fact table
- DM comprises of a set of dimensional tables
- All of the given option Page 103

44) ----- can be used when some columns are rarely accessed rather than other columns or when the table has wide rows or header or both:

- Horizontal splitting
- Pre-joining
- Vertical splitting page 56
- Derived attributes

45) Which of the following is an example of derived attributes?

- Age page 61
- Size
- Color
- Length

46) The online high performance transaction processing was evolved in -----:

- 1980
- 1975 page 12

- 1977

- 1965

47) Cube is a logical entity containing values of a certain aggregation level at an intersection of a combination of ----- :

- Facts

- Dimension page 88

- Summary tables

- Primary and foreign key

48) Which of the following is TRUE regarding Entity relationship modeling?

- It does not really model business, but models the micro relationships among data elements.

- ER modeling does not have “business rules,” it has “data rules

- ER modeling helps retrieval of individual records having certain critical identifiers.

- All of the given option page 102

49) ----- facilitates a mobile computing paradigm:

- DOLAP page 78

- HOLAP

- ROLAP

- MOLAP

50) The main reason(s )for the increase in cube size may be:

- Increase in the number of dimensions

- Increase in the cardinality of the dimensions

- Increase in the amount of detail data

- All of the given options page 87

51) Suppose the amount of data recorded in an organization is doubled in year. This increase in -----  
-----:

- Linear

- Quadratic

- Exponential page 15

- Logarithmic

52) The data in the data warehouse is ----- :

- Volatile

- Non-volatile page 69

- Static

- Non-structured

53) ----- models the macro relationships among data elements with an overall deterministic strategy:

- Dimensional model page 102

- Entity relationship model

- Object oriented model
- Structured model

54) De- normalization affects:

- Database size and query performance page 52
- Database usability and query reliability
- Database availability and query success
- None of the given options

55) ----- technique requires a separate column to specify the time and date when the last modification was occurred:

- Checkmarks
- Timestamps page 150
- Just-in-Time
- Real Time extraction

56) Which of the de-normalization technique squeezes master table into detail?

• Pre-joining page 58

- Horizontal splitting
- Vertical splitting
- Adding redundant column

57) De-normalization can help:

- Minimize joins
- Minimize foreign keys
- Resolve aggregates
- All of the given options page 51

58) The domain of the “gender” field in some database may be (‘F’,’M’) or as (“Female”, “Male”) or even as (1, 0). This is:

- Primary key problem

• Non primary key problem page 163

- Normalization problem
- All of the given option

59) Increasing level of normalization ----- number of tables:

• Increases page 51

- Decreases
- Does not effect
- None of the given option

60) Which of the following is not a Data Quality Validation Technique:

- Referential integrity
- Using Data Quality Rules
- Data Histograming
- Indexes page 189

61) This technique can be used when column from one table is frequently accessed in a large scale

join in conjunction with a column from another table:

- Horizontal splitting
- Pre-joining
- Adding redundant column page 58
- Derived attributes

62) Data cleansing requires involvement of domain expert because:

- Domain expert has deep knowledge of data aggregation
- Change Data captures requires involvement of domain expert
- Domain knowledge is required to correct anomalies page 158
- Domain expert has deep knowledge of data summarization

63) Relational databases allow you to navigate the data in ----- that is appropriate using the

primary , foreign key structure with in the data model:

- Only One Direction
- Any Direction page 19
- Two Direction
- None of these

64) History is excellent predictor of the -----:

- Past
- Present
- Future page 15
- History

65) De- normalization is the process of selectively transforming normalized relations into un

normalized physical record specifications , with the aim to:• Well structure the data

- Well model the data
- Reduce query processing time page 50
- None of the given option

66) ----- gives total view of an organization:

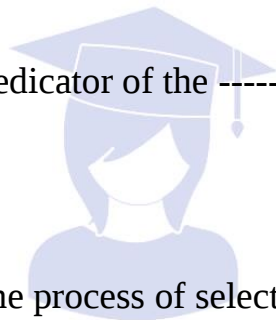
- OLAP
- OLTP
- Data Warehouse page 16
- Database

67) Suppose in system A, the possible values of “Gender” attribute were “Male”& “Female”,

however in data warehouse ,the values stored were “M” for male and “F” for female. This above

scenario is an example of :

- One-to-one scalar transformation page 144
- One-to-many element transformation



- Many-to-one element transformation
- Many-to-many element transformation

68) Enrichment is one of the basic tasks in data ----- :

- Extraction
- Transformation page 138
- Loading
- Summarization

69) Which of the following is not a technique of De-normalization?

- Pre-joining
- Splitting tables
- Adding redundant columns

70) Which of the following is an example of Additive Facts?

- Sales Amount page 119
- Average
- Discount
- Ratios

71) Robotic libraries are needed for -----:

- Cubes
- Data marts
- Data warehouse page 131
- Aggregates

72) Normally ROLAP is implemented using -----

• **Star schema page 87**

- Hybrid schema
- Pre-defined aggregate
- All of the given options

73) The relation R will be in 2nd Normal Form if

- It is in 1NF and each cell contains single value
- It is in 1NF and each non key attribute is dependent upon entire primary key page 44

• It is in 1NF and non key attribute is dependent upon a single column of composite primary key

74) In ----- nested loop join of quadratic time complexity does not hurt the performance

- Typical OLTP environments page 22
- Data warehouse
- DSS
- OLAP

75) In Extract, Load, Transform(ELT) process, data transformation -----:-----:

- Takes place on the data warehouse server page 147

- Takes place on a separate transformation server
- Depends on the nature of the source database
- Does not take place

76) Node of a B-Tree is stored in memory block and traversing a B-Tree involves ----- page

faults:

- $O(n \log n)$
- $O(\log n)$  page 22
- $O(n)$
- $O(n^2)$

77) As dimensions get less detailed (e.g. , year vs. day) cubes get -----

- Smaller page 84
- Larger
- Partitioned
- Merged

78) Which of the following is not a technique of “ Changed Data Capture” in currently used Modren Source System?

- Timestamps
  - Partitioning
  - Triggers
  - Dimensional Modeling page 150
- 79) The trade-offs of de-normalization is/are:

- Storage
- Performance
- Ease-of-use
- All of the given options page 62

80) If actual data structure does not conform to documented formats then it is called:

- Syntactically dirty data page 160
- Semantically dirty data
- Coverage anomaly
- Extraction issue

81) “Header size is reduced, allowing more rows per back , thus reducing I/O” .The above statement is TRUE with respect to:

- Vertical splitting page 56
  - Horizontal splitting
  - Adding redundant column
  - None of the given options
- 82) ----- Breaks a table into multiple tables based upon common column values
- Horizontal splitting page 54

- Vertical splitting
- Adding redundant column
- None of the given option

83) Which of the following is NOT an example of derived attribute?

- Age
- CGPA
- Area of rectangle
- Height (Conceptual)

84) Which of the following is NOT an example of derived attribute?

- Age
- CGPA
- Annual Salary
- Email Address (Conceptual)

85) If a table is expected to have six columns but some or all of the records do not have six columns then it is example of:

- Syntactically dirty data page 160
- Semantically dirty data
- Coverage anomaly
- Extraction issue

86) MDX by Microsoft is an example of -----:

- HOLAP
- DOLAP
- ROLAP
- None of the given options page 79

87) The growth of master files and magnetic tapes exploded around the mid- -----

- 1950s
- 1960s page 12
- 1970s
- 1980s

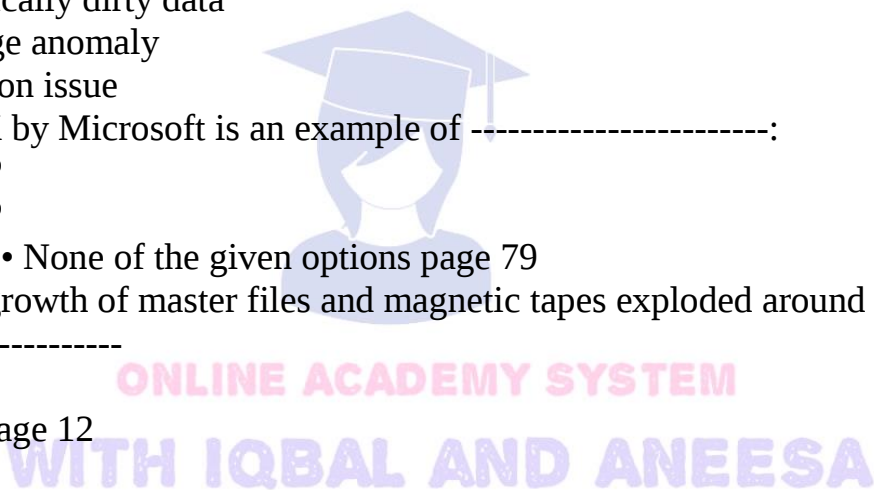
88) If one or more records in a relational table do not satisfy one or more integrity constraint , then the data:

- Is syntactically dirty
- Is semantically dirty page 160
- Has Coverage anomaly
- Has extraction issue

89) OLAP is:

- Analytical processing page 69
- Transaction processing
- Additive processing
- Active processing

90) One of the possible issues faced by web scrapping is that:



- Web pages may contain junk data page 141
- Web pages do not contain multiple facts
- Web pages do not contain multiple dimensions
- Web pages does not support transformation

91) Which of the following is\are example of dimension:

- Product page 79
- Region• Data
- None of the given

92) An OLTP system is always good at -----:

- Evolving data page 122
- Keeping static data
- Tracking past data
- Maintaining historic data

93) In case of multiple sources for the same data element , we need to prioritize the source systems

per element based, the process is called:

- Ranking page 143
- Prioritization
- Element selection
- Measurement selection

94) One feature of Change Data Capture (CDC) is that:

- It pre-calculates changed aggregates
- It loads the transformed data in real time
- It only processes the data has been changed
- It can automate the transformation of extracted data page 149

95) In ----- SQL generation is vastly simplified for front-end tools when the data is highly structure:

- MOLAP
- Star Schema page 107
- Hybrid schema
- Object oriented schema

96) Dirty data means:

- Data cannot be aggregated
- Data contains non-additive facts
- Data does not fulfill dimensional modeling rules• Data does not conform to proper domain definitions page 158

97) In Context of Change Data Capture (CDC) sometimes a ----- object can be used to store recently modified data:

- Buffer table
- Change table page 149
- Checkmark table

- Change control table

98) “Sometimes during data collection complete entities are missed”.

This statement is an example

of :

- Missing tuple page 161
- Missing attribute
- Missing aggregates
- Semantically dirty data

99) Table collapsing technique is applied in case of:

- One-by-one relation or many-to –many relation page 52
- One-to-many relation
- Many-to-many relation
- None of the given option

100)

Which of the following is an example of dimension?

- Product
- Region
- Date
- All of the given option page 78

101)

Data warehouse stores -----:

- Operational data• Historical data page 24
- Meta data
- Log files data

102)

The business process covered by ER diagrams:

- Do not co-exist in time and space page 109
- Co-exist in time and space
- Do not physically exist in real time context
- None of the given options

103)

The main goal of normalization is to eliminate -----:

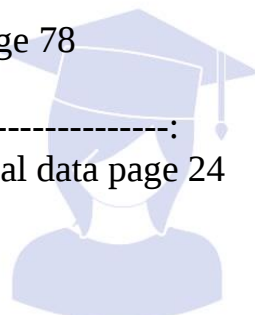
- Data redundancy page 41
- Data sharing
- Data security
- Data consistency

104)

Serious ---- involves decomposing and reassembling the data:

- Data cleansing page 168
- Data transformation
- Data loading
- Data extraction

105)



In the data warehouse environment the data is -----

- Subject- oriented page 69
- Time- oriented
- Both subject and time oriented
- Neither time-oriented nor subject- oriented

106)

For large record spaces and large number of records , the run time of the clustering

algorithms:

- Prohibitive page 164
- Static• Exponential
- Numerical

107)

----- can result in costly errors, such as , False frequency

distributions and incorrect

aggregates due to double counting:

- Data duplication page 165
- Data reduction
- Data anomaly
- Data transformation

108)

The degree to which values are present in the attributes that require them is known as -

-----:

- Completeness page 185
- Uniqueness
- Accessibility
- Consistency

109)

Time complexity of Key Creation process in basic Sorted Neighborhood (BSN) Method is

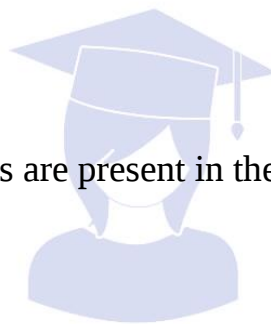
-----:

- $O(n \log n)$
- $O(\log n)$
- $O(n)$  page 171
- $O(2n)$

110)

Which of the following is an example of slowly changing dimensions?

- Inheritance page 124
- Aggregation
- Association
- Asset disposal111)



ONLINE ACADEMY SYSTEM  
WITH IQBAL AND ANEESA

The ----- operator proves useful in more complex metrics applicable to the dimensions and accessibility:

- Max page 188
- Min
- Max and Min
- None of the given

112)

In OLAP, the typical write operation is ----- :

• Bulk insertion page 75

- Single insertion
- Sequential insertion
- No insertion

113)

The issue(s) of “ Adding redundant column” includes(s):

- Increase in table size
- Maintenance
- Loss of information
- All of the given option page 65

114)

----- is applicable in Profitability analysis:

- OLTP
- Data warehouse page 36,37
- Information System(IS)
- Management Information System(MIS)115)

The hardware (CPU) utilization in data warehouse environment is full or ----- :

- Fixed
- Partial
- Not at all page 24
- Slow

116) Time variant is a characteristics of data warehouse which means:

- Data loaded in data warehouse will be time stamped page 20
- Data can be loaded in data warehouse anytime
- Data can be loaded in data warehouse only at a particular time
- Data cannot be loaded in data warehouse with respect to time

117)

In which class of aggregates AVERAGE function can be placed:

- Algebraic page 120
- Distributed
- Associative
- Holistic

118) Considered the following Employee table and identify the column which causes that the table is not in first normal form(1NF):

Employee(Emp\_ID, Emp\_Name ,Emp\_skills, Emp\_Designation)

- Emp\_ID
- Emp\_Name
- Emp\_skills page 43(conceptual)• Emp\_Designation

119)

The application of data and information leads to -----

- Intelligence
- Experience
- Knowledge page 11
- Power

120)

----- segregate data into separate partitions so that queries do not need to

examine all data in a table when WHERE clause filters specify only a subset of the partitions.

- Pre-joining technique
- Collapsing table technique
- Horizontal splitting technique page 56
- Vertical splitting technique

1. Which logical data extraction has significant performance impacts on the data warehouse server? Incremental extraction
2. ....is the lowest level of detail or the atomic level of data stored in the warehouse? Grain
3. ....incorporates the concept of product quality process control quality assurance and quality improvement? Realistic data quality management
4. Which is a/an step of data cleansing procedure? Aggregation
5. Two interesting examples of quality dimensions that can make use of the min operator are.....? reliability and appropriate amount of data
6. The degree of similarity between two records often measured by a numerical value between.....? 0 and 1
7. Which of the following is an example of non additive facts?Discount in percentage
8. Which one among the following is a de normalization technique?Adding redundant columns
- 9.