CS-614 (DATA Warehousing)

Grand Quiz Dec 2020

Prepared By: MCS of Virtuallians

If you found any error in the file, Please correct it accordingly.

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 for (1NF) if it does not contain_____.
    a. Single value column
    b. Multivalve column google
    c. Derived Column
    d. Composite Column

MCS of Virtuallians Group Team

14. In a fully normalized database, too many_____ are required.
    a. Values
    b. Joins Page - 49
    c. Queries
    d. Conditions

MCS of Virtuallians Group Team

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 studding 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

MCS of Virtuallians Group Team

17. OLTP implementations are fully _____.
    a. Normalized Page-30
    b. Demoralized
    c. Predictive
    d. Additive

MCS of Virtuallians Group Team

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

MCS of Virtuallians Group Team

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

a. Aggregation Page - 135

b. Selection

c. Splitting / joining

d. Conversion

MCS of Virtuallians Group Team

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

MCS of Virtuallians Group Team

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 takes _____

a. O(n) page 190

b. O(1)

c. O(log n)

d. None of the given

MCS of Virtuallians Group Team

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.

MCS of Virtuallians Group Team

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

MCS of Virtuallians Group Team

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
    d. 1990s

MCS of Virtuallians Group Team

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
    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

MCS of Virtuallians Group Team

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 page144
    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 page153
    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 page85

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 page103

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

MCS of Virtuallians Group Team

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

MCS of Virtuallians Group Team

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

MCS of Virtuallians Group Team

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

MCSof Virtuallians Group Team

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 page50
    d. None of the given options

MCSof Virtuallians Group Team

81. Which of the following is not a task of Data Transformation?
    a. Conversion
    b. Summarization
    c. Enrichment
    d. Full Data Refresh page135, 154

MCSof Virtuallians Group Team

82. _____ gives total view of an organization.
    a. OLAP
    b. OLTP
    c. Data Warehouse page15, 16
    d. Database

MCSof Virtuallians Group Team

83. Enrichment is one of the basic tasks in data _____
    a. Extraction
    b. Transformation page138
    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 page50, 52

MCSof Virtuallians Group Team

85. Which of the following is an example of Additive Facts?
    a. Sales Amount page119
    b. Average
    c. Discount
    d. Ratios
86. Robotic libraries are needed for _____
    a. Cubes
    b. Data marts
    c. Data warehouse page131
    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. (n lg n)
    b. O (Ig n) page22
    c. O(n)
    d. (n2)

MCSE Virtuallians Group Team

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

MCS of Virtuallians Group Team

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

a. Mid-1960 page 12

b. Mid-1970

c. Mid-1980

d. Mid-1990

MCSof Virtuallians Group Team

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

MCSof Virtuallians Group Team

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

MCSof Virtuallians Group Team

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

MCSof Virtuallians Group Team

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:

MCSof Virtuallians Group Team

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

MCSof Virtuallians Group Team

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

MCSof Virtuallians Group Team

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

MCSE Virtuallians Group Team

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

MCSof Virtuallians Group Team

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

MCSof Virtuallians Group Team

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?

MCSof Virtuallians Group Team

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?

MCSof Virtuallians Group Team

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 we 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

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

MCSof Virtuallians Group Team

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 _____, differe3nt 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 3nd 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

MCSE Virtuallians Group Team

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. Inteligence
   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 → Auther _Name, Auther _Age
   Auther_Name →Auth0r_Age
   a. Partial dependency
   b. Transitive dependency page 173 form CS403
   c. Full-functional dependency

MCS of Virtuallians Group Team

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)

MCS of Virtuallians Group Team

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

a. Large
b. Fixed
c. Moderate
d. Small page 22

MCS of Virtuallians Group Team

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

MCS of Virtuallians Group Team

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

MCS of Virtuallians Group Team

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

MCS of Virtuallians Group Team

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 ware house.

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

MCSet Virtuallians Group Team

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

MCS of Virtuallians Group Team

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

MCS of Virtuallians Group Team

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?

MCS of Virtuallians Group Team

   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 __

MCS of Virtuallians Group Team

   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

MCS of Virtuallians Group Team

   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
   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
   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. Date 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

MCS of Virtuallians Group Team

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 page103
   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

MCSE Virtuallians Group Team

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.

MCS of Virtuallians Group Team
MCS of Virtuallians Group Team
MCS of Virtuallians Group Team
MCS of Virtuallians Group Team


FAMILY GROUP
ARSL SHANI &
AINA MALIK
Mcs Family Group
Created by Arsl, 3/30/20
@nameshadowart

CS614 GRAND QUIZ

SOLUTION MADE BY:

ARSL SHANI, AINA MALIK & MCS FAMILY GROUP TEAM

ARSL SHANI, AINA MALIK & MCS FAMILY GROUP

1) Online Extraction is a kind of ---------- data extraction:
   - Physical page 132
   - Logical
   - Dimensional
   - Multi valued

   - **Physical Extraction**
     - Online Extraction

2) The ---------- saw the advent of disk storage, or DASD (direct Access Storage Device) :
   - 1960s
   - 1970s page 13
   - 1950s
   - 1990s

   The 1970s saw the advent of disk storage, or DASD (Direct Access Storage Device).

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

   It is very difficult to extract the data from different source systems, because by definition they are heterogeneous, and as a consequence, there are multiple representations of the

4) Which of the following is not a task of Data Transformation?
   - Conversion
   - Summarization
   - Enrichment
   - Full Data Refresh page 135

16.6 Data Transformation

- **Basic tasks**
  - Selection
  - Splitting/Joining
  - 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

21.3 Orr's Laws of Data Quality

Law #1: "Data that is not used cannot be correct!"

© Copyright Virtual University of Pakistan

Law #2: "Data quality is a function of its use, not its collection!"

Law #3: "Data will be no better than its most stringent use!"

Law #4: "Data quality problems increase with the age of the system!"

Law #5: "The less likely something is to occur, the more traumatic it will be when it happens!"

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

- Online
- Offline page 134
- Incremental
- Full

ARSL SHAN MILY GROUP

- **Offline Extraction**
  - Data NOT extracted directly from the source system, instead staged explicitly outside the original source system.
  - Data is either already structured or was created by an extraction routine.
  - Some of the prevalent structures are:
    - Flat files

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

Some of the advantages of CDC are;

1. Database extraction from INSERT, UPDATE, and DELETE operations occurs immediately, at the same time the changes occur to the source tables.

2. In the absence of CDC database extraction is marginal at best for INSERT operations, and problematic for UPDATE and DELETE operations, because the data is no longer in the table.

3. Stages data directly to relational tables; there is no need to use flat files. In the absence of CDC, the entire contents of tables are moved into flat files.

4. Log tape processing requires no extra I/O in the middle of online processing, as does

8) The most common range partitioning is on
   - Color
   - Date page 66
   - PhoneNo
   - Name

The most common use of range partitioning is on date. This is especially true in data

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
   - Composite column

Although the table is shown in 1NF i.e. it contains atomic values, there are no repeating values, there is no aggregation, yet it contains redundant data. For example, information

ARSL SHAN ALI A MAL ILY GROUP

A table is in the first normal form (1NF) if

- The table has a primary key.
- No single column has multiple values.
- The nonprimary key columns depend on the primary key.

10) In a fully normalized database, too many _____are required
    - Values
    - Joins page 49
    - Queries
    - Conditions

There should be a balance between normalized and de-normalized forms. In a fully normalized form, too many joins are required and in a totally de-normalized form, we

11) In the data warehouse, data is collection from ---------- sources:
    - Homogeneous
    - Heterogeneous page 21
    - External
    - Internal

into a box and soon you have a Data Warehouse up and running. A Data Warehouse evolves over time, you don't buy it. Basically it is about taking/collecting data from different heterogeneous sources. Heterogeneous means not only the operating system is

12) De-normalization is more like a "controlled crash" with the aim to ---------- without loss of information:
    - Check
    - Balance
    - Decrease
    - Enhance page 49

▪ It is not chaos, more like a "controlled crash" with the aim of performance enhancement without loss of information.

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

Quality improvement is making all efforts to increase effectiveness and efficiency in meeting accepted customer expectations.

14) ---------- is the application of intelligence and experience to get common goals.

ARSL SHANTANA MALIK & MOEEZ FAMILY GROUP

- Wisdom page 11
- Education
- Power
- Information

Intelligence is appreciation of "why", and finally wisdom (not shown in the figure-1.1) is the application of intelligence and experience toward the attainment of common goals, and wise people are powerful. Remember knowledge is power.

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

- Aggregation
- Enrichment page 136
- Splitting joining
- Conversion

Enrichment: This task is the rearrangement and simplification of individual fields to make them more useful for the data warehouse environment. You may use one or more

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

Step-2: Grain of a fact table = the meaning of one fact table row. Determines the maximum level of detail of the warehouse.

17) Normalization ---------- :

- Reduces redundancy page 41
- Increases redundancy
- Reduces joins
- Reduces tables

Normalization is the process of efficiently organizing data in a database by decomposing (splitting) a relational table into smaller tables by projection. There are basically two goals of normalization as follows:

1. Eliminate redundant data (for example, storing the same data in more than one table)

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

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

ARSL SHAN MALIK & MCS FAMILY GROUP

- Typical grains
  - Individual Transactions
  - Daily aggregates (snapshots)
  - Monthly aggregates

19) Multi-dimensional databases(MDDs) typically use ---------- formats to store pre-summarized cube structures:
- SQL
- Proprietary file page 79
- Object oriented
- Non-proprietary file

Multi-dimensional databases (MDDs) typically use proprietary file formats to store pre-summarized cube structures. There are dozens of vendors who provide MDD products.

20) ---------- provides a combination of "relational databases access" and "cube" data structures within a single framework:
- HOLAP page 78
- DOLAP
- MOLAP
- ROLAP

HOLAP provides a combination of relational database access and "cube" data structures within a single framework. The goal is to get the best of both MOLAP and ROLAP:

21) Data Warehouse provides the best support for analysis while OLAP carries out the ---------- task:
- Mandatory
- Whole
- Analysis page 69
- Prediction

Data Warehouse provides the best support for analysis while OLAP carries out the analysis task. Although there is more to OLAP technology than data warehousing, the

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

Vertical splitting 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.

23) OLAP implementations are highly/completely ---------- :
- Normalized
- Demoralized page 69

ARSL SHAHZAD MALIK & MCS FAMILY GROUP

- Predictive
- Additive

OLAP implementations are either highly de-normalized or completely de-normalized,

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

Although the table is shown in 1NF i.e. it contains atomic values, there are no repeating values, there is no aggregation, yet it contains redundant data. For example, information

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

are used to capture the relationships between dimensions. Fact less fact table captures the many-to-many relationships between dimensions, but contains no numeric or textual

26) Which of the following is not an example of dimension:
- Product
- Date
- Region
- Sales volume page 78
  - Dimensions are key business factors for analysis:
    - Geographies (city, district, division, province,...)
    - Products (item, product category, product department,...)
    - Dates (day, week, month, quarter, year,...)

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

Many people consider business processes as Data marts i.e. in their view organizational or departmental function is referred as the business process. That is why such people criticize DM as being a data mart oriented approach. However, in Kimball's view, it is a

28) In a fully normalized form:
- To many joins are required page 49
- Relationships lose their significance
- No joins are required

ARSL SHAHZAD, ARSLIK & MCS FAMILY GROUP

- Data integrity becomes an issue

There should be a balance between normalized and de-normalized forms. In a fully normalized form, too many joins are required and in a totally de-normalized form, we

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

- **Non-additive facts:**
  - **Averages (average sales price, unit price)**
  - **Percentages (% discount)**

30) Which of the following is not a CUBE operation?
- ANSI SQL page 81
- Roll Up
- Drill Down
- Pivoting

stored in a cube. There are four fundamental cubes operations which are (i) rollup (ii) drill down (iii) slice and dice and (iv) pivoting. In the rollup operation the level of aggregation

31) ---------- allows download of "cube" structures to a desktop platform without the need for shared or cube server:
- MPLAP
- ROLAP
- DOLAP page 78
- HOLAP

DOLAP allows download of "cube" structures to a desktop platform without the need for shared relational or cube server. This model facilitates a mobile computing paradigm.

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

ROLAP or a Relational OLAP provides access to information via a relational database using ANSI standard SQL.

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

ARSL SHAN KHAN MALIK & MCS FAMILY GROUP

- Expression indexing
- None of the given option

Expression partitioning 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. For example, partitioning a table by expression

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

Taken jointly, the extract programs or naturally evolving systems formed a spider web, also called "legacy systems" 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

data in different formats and focus on different aspects of the subject. The data has to be checked, cleansed and transformed into a unified format to allow easy and fast access.

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

One-to-many element transformations
- 4 x 20 address field → House/Flat, Road/Street, Area/Sector, City.

element form the source system results in several columns in the DW. Consider the 6×30 address field (6 lines of 30 characters each), the requirement is to parse it into street address lines 1 and 2, city, sate and zip code by applying a parsing algorithm.

37) In ~~decision support~~ system ease of use in achieved by:
- Normalization
- Denormalization page no 49
- Drill up
- Drill down

Normalization is a rule of thumb in DBMS, but in DSS ease of use is achieved by way of denormalization.

ARSL SHAN[illegible] MALL & MCS FAMILY GROUP

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

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

13.5 How to simplify an ER data model?

- Two general methods:
  - De-Normalization
  - Dimensional Modeling (DM)

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

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

the data into the DWH. There is a significant part after data transformation that involves data cleansing i.e. there is a C also here. However, ETL is a pretty standard definition,

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

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

19.3 Syntactically Dirty Data

Lexical errors: For example, assume the data to be stored in table form with each row representing a tuple and each column an attribute. If we expect the table to have five columns because each tuple has five attributes but some or all of the rows contain only four columns then the actual structure of the data does not conform to the specified format.

Irregularities are concerned with the non-uniform use of values, units and abbreviations.

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

Format Revisions: You will come across these quite often. These revisions include changes to the data types and lengths of individual fields 12 digit new NID. In your

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

ARSL SHAN MALIK & MCS FA GROUP

- Flexibility
- Maintainability
- Security
- Scalability page 85
  - The divide&conquer cube partitioning approach helps alleviate the scalability limitations of MOLAP implementation.

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
  - Inherently dimensional in nature, with a single central fact table and a set of smaller dimensional tables.

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

Vertical splitting can be used when some columns are rarely accessed rather than other columns or when the table has wide rows or header or both. Thus the infrequently

45) Which of the following is an example of derived attributes?
- Age page 61
- Size
- Color
- Length

Age is also a derived attribute, calculated as Current_Date – DoB (calculated periodically).

46) The online high performance transaction processing was evolved in ----------:
- 1980
- 1975 page 12
- 1977
- 1965

1975

Online high performance transaction processing

47) Cube is a logical entity containing values of a certain aggregation level at an intersection of a combination of ---------- :
- Facts
- Dimension page 88

ARSL SHAMIL & MCS FAMILY GROUP

- Summary tables
- Primary and foreign key

  - **Cube is a logical entity containing values of a certain fact at a certain aggregation level at an intersection of a combination of dimensions.**

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

1- ER models are constituted to (a) remove redundancy from the data model, (b) facilitate retrieval of individual records having certain critical identifiers, and (c) therefore, optimize On-line Transaction Processing (OLTP) performance.

2- ER modeling does not really model a business; rather, it models the micro relationships among data elements. ER modeling does not have "business rules," it has "data rules.“

49) ---------- facilitates a mobile computing paradigm:
- DOLAP page 78
- HOLAP
- ROLAP
- MOLAP

DOLAP allows download of “cube” structures to a desktop platform without the need for shared relational or cube server. This model facilitates a mobile computing paradigm.

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

size increases. There may be many reasons for the increase in cube size, such as increase in the number of dimensions, or increase in the cardinality of the dimensions, or increase in the amount of detail data or a combination of some or all these aspects. Thus there is an

51) Suppose the amount of data recorded in an organization is doubled in year. This increase in ------ ----
- Linear
- Quadratic
- Exponential page 15
- Logarithmic

contents carried. On the practical side, it has been noted that the amount of data recorded in an organization doubles every year and this is an exponential increase.

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

ARSL SHAHZAD MAHMOOD & MOHSIN FAMILY GROUP

- Non-volatile page 69
- Static
- Non-structured

A data warehouse is a "subject-oriented, integrated, time varying, non-volatile collection of data that is used primarily in organizational decision making. Typically, the data

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

DM
Constituted to optimize DSS query performance.
Models the macro relationships among data elements with an overall deterministic strategy.

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

De-normalization will ultimately affect the database size and query performance.

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

Timestamps
The tables in some operational systems have timestamp columns. The timestamp specifies the time and date that a given row was last modified. If the tables in an operational system

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

- Pre-joining page 58
- Horizontal splitting
- Vertical splitting
- Adding redundant column

ARSL SHAN AMALIK & MCS FAMILY GROUP

To avoid the run-time join, we use the pre-join technique and "squeeze" the sales master information into the detail table. The obvious drawback is repetition of facts from the

57) De-normalization can help:
- Minimize joins
- Minimize foreign keys
- Resolve aggregates
- All of the given options page 51

De-normalization can help minimize joins and foreign keys and help resolve aggregates.

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

Non primary key problems…

1. Different encoding in different sources.
2. Multiple ways to represent the same information.
3. Sources might contain invalid data.
4. Two fields with different data but same name.

1. Data may be encoded differently in different sources. The domain of a "gender" field in some database may be {'F', 'M'} or as {"Female", "Male"} or even as {1, 0}.

59) Increasing level of normalization ------------- number of tables:
- Increases page 51
- Decreases
- Does not effect
- None of the given option

The higher the level of normalization, the greater will be the number of tables in the DSS

60) Which of the following is not a Data Quality Validation Technique:
- Referential integrity
- Using Data Quality Rules
- Data Histograming
- Indexes page 189

ARSL SHAHZAD ANWAR'S FAMILY GROUP

22.2 Data Quality Validation Techniques

- Referential Integrity (RI).
- Attribute domain.
- Using Data Quality Rules.
- Data Histograming.

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

8.3 Adding Redundant Columns

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. To avoid this join, the

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

Usually the process of data cleansing can not be performed without the involvement of a domain expert because the detection and correction of anomalies requires detailed domain knowledge. Data cleansing is therefore described as semi-automatic but it should be as

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

warehouse, because relational databases allow you to navigate the data in any direction that is appropriate using the primary, foreign key structure within the data model.

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

- Past
- Present
- Future page 15
- History

- History is excellent predictor of the future.

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

ARSL SHAN ALI MEMON FAMILY GROUP

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

information will not be lost. De-normalization is the process of selectively transforming normalized relations into un-normalized physical record specifications, with the aim of reducing query processing time. Another fundamental purpose of denormalization is to

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

- OLAP
- OLTP
- Data Warehouse page 16
- Database

So why would you want data warehouse in your organization? First of all a data warehouse gives a total view of an organization. If you look at the operational system i.e.

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

Simple scalar transformation is a one-to-one mapping from one set of values to another set of values using straightforward rules. For example, if 0/1 corresponds to male/female in one source system, then the gender is stored as male/female in the DW.

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

- Extraction
- Transformation page 138
- Loading
- Summarization

Enrichment is one of the basic tasks of data transformation

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

- Pre-joining
- Splitting tables
- Adding redundant columns
- ER modeling page 52

ARSL SHAHZAD ALI & MCS FAMILY GROUP

Now we will discuss de-normalization techniques that have been commonly adopted by experienced database designers. These techniques can be classified into four prevalent strategies for denormalization which are:

1. Collapsing Tables.
   - Two entities with a One-to-One relationship.
   - Two entities with a Many-to-Many relationship.

2. Pre-joining.

3. Splitting Tables (Horizontal/Vertical Splitting).

4. Adding Redundant Columns (Reference Data).

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

- Sales Amount page 119
- Average
- Discount
- Ratios

There can be two types of facts i.e. additive and non-additive. Additive facts are those facts which give the correct result by an addition operation. Examples of such facts could be number of items sold, sales amount etc. Non-additive facts can also be added, but the

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

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

use a robotic tape library and all the associated setup. Typically you don't need robotic libraries for the cubes and data marts, but you do need them for data warehouse. But

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

- Star schema page 87
- Hybrid schema
- Pre-defined aggregate
- All of the given options

OLAP. ROLAP tools will query the relational database using SQL generated to conform to a framework using the facts and dimensions paradigm using the star schema.

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
- It is in 1NF and Primary key is composite

A relational table is in second normal form 2NF if it is in 1NF and every non-key column is fully dependent upon the primary key. That is, every non-key column must be

ARSL SHAHZAD & ALI KHAN & ITS FAMILY GROUP

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

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

In typical OLTP environments, the size of tables are relatively small, and the rows of interest are also very small, as the queries are confined to specifics. Hence traditional joins such as nested-loop join of quadratic time complexity does not hurt the performance

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

ELT: Extract, Load, Transform in which data transformation takes place on the data warehouse server.

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(n2)

Trees becomes very large. Typically the node of a B-Tree is stored in a memory block, and traversing a B-Tree involves O(log n) page faults. This is highly undesirable, because

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

- Smaller page 84
- Larger
- Partitioned
- Merged

Storage issue: As dimensions get less detailed (e.g., year vs. day) cubes get much smaller, but storage consequences for building hundreds of cubes can be

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

ARSL SHAHZAD MALIK & MUHAMMAD FAMILY GROUP

Several techniques are used for implementing changed data capture in currently used state of the art source systems:

- Timestamps
- Partitioning
- Triggers

79) The trade-offs of de-normalization is/are:

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

The trade-offs of denormalization are as follows:

- Storage
- Performance
- Ease-of-use
- Maintenance

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

19.3 Syntactically Dirty Data

Lexical errors: For example, assume the data to be stored in table form with each row representing a tuple and each column an attribute. If we expect the table to have five columns because each tuple has five attributes but some or all of the rows contain only four columns then the actual structure of the data does not conform to the specified format.

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

Splitting Tables: Vertical Splitting

- Splitting and distributing into separate files with repeating primary key.
- Infrequently accessed columns become extra "baggage" thus degrading performance.
- Very useful for rarely accessed large text columns with large headers.
- Header size is reduced, allowing more rows per block, thus reducing I/O.

ARSL SHANIA & MCS LY GROUP

82) ---------- Breaks a table into multiple tables based upon common column values

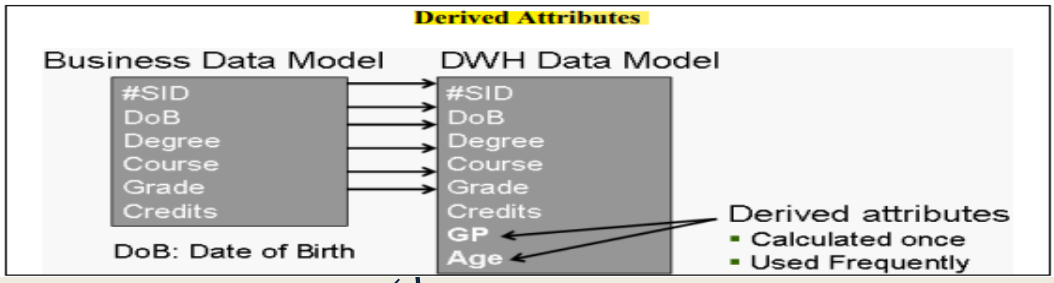
- Horizontal splitting page 54
- Vertical splitting
- Adding redundant column
- None of the given option

Splitting Tables: Horizontal splitting

Breaks a table into multiple tables based upon common column values. Example: Campus

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

19.3 Syntactically Dirty Data

Lexical errors: For example, assume the data to be stored in table form with each row representing a tuple and each column an attribute. If we expect the table to have five columns because each tuple has five attributes but some or all of the rows contain only four columns then the actual structure of the data does not conform to the specified format

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

- HOLAP
- DOLAP
- ROLAP

ARSL SHAMLA MALIK FAMILY GROUP

- None of the given options page 79

MOLAP Implementations

- No standard query language for querying MOLAP
  - *No SQL !*
- Vendors provide proprietary languages allowing business users to create queries that involve pivots, drilling down, or rolling up.
  - E.g. MDX of Microsoft

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

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

Around the mid-1960s, the growth of master files and magnetic tapes exploded. Soon

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

Semantically dirty data

Integrity constraint violations describe tuples (or sets of tuples) that do not satisfy one or more of the integrity constraints. Integrity constraints are used to describe our

89) OLAP is:

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

OLAP is Analytical Processing instead of Transaction Processing. It is also NOT a

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

17.3 Web scrapping

- Lot of data in a web page, but is mixed with a lot of "junk".

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

- Product page 79
- Region

ARSL SHAHZAD NAWAZ & MCS FAMILY GROUP

- Data
- None of the given

WHERE clause is executed. Some examples of dimension are geography, time, products etc. Note that typically dimensions have hierarchies, for example geography dimension

92) An OLTP system is always good at ----------:
- Evolving data page 122
- Keeping static data
- Tracking past data
- Maintaining historic data

discussed before. A good OLTP system is always evolving. Orders are being filled and,

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

This means you need to establish ranking in the source system on per element (attribute) basis. Ranking is all about selecting the "right" source system. Rank establishment has to

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

from one or more source databases, into the data warehouse for analysis. Change Data Capture (CDC) quickly identifies and processes only the data that has changed, not entire tables, and makes the change data available for further use.

95) In ---------- SQL generation in vastly simplified for front-end tools when the data is highly structure:
- MOLAP
- Star Schema page 107
- Hybrid schema
- Object oriented schema

The ultimate goal of a star schema design is to put into place a physical data model capable of very high performance to support iterative analysis adhering to an OLAP model of data delivery. Moreover, SQL generation is vastly simplified for front-end tools when the data is highly structured in this way.

96) Dirty data means:
- Data cannot be aggregated
- Data contains non-additive facts
- Data does not fulfill dimensional modeling rules

ARSL SHAHZAIB & ARSLAN FAMILY GROUP

- Data does not conform to proper domain definitions page 158

and you don't even want to consider what work cleaning it up would take. By "dirty," I mean that it does not conform to proper domain definitions or "make sense." The age-old

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

user tables is recorded. For some environments, the change data is then stored in a database object called a *change table*, and the changed data is made available to applications in a controlled way.

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

Missing tuples
Result from omissions of complete entities existent in the mini-world that are not represented by tuples in the data collection. Anomalies could further be classified

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

1. Collapsing Tables.
   - Two entities with a One-to-One relationship.
   - Two entities with a Many-to-Many relationship.

100) Which of the following is an example of dimension?

- Product
- Region
- Date
- All of the given option page 78
  - Dimensions are key business factors for analysis:
    - Geographies (city, district, division, province,...)
    - Products (item, product category, product department,...)
    - Dates (day, week, month, quarter, year,...)

101) Data warehouse stores ----------:

- Operational data

ARSL SHAHZAD MALIK 21 MCS FACULTY GROUP

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

customers and old transactions. In the data warehouse we save the historical 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

A typical ER diagram covers the entire spectrum of the business, actually covers every possible business process. However, in reality those multitudes of process do not co-exist in time and space (tables). As a consequence, an ER diagram is overly complex, and is a

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

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

Normalization is the process of efficiently organizing data in a database by decomposing (splitting) a relational table into smaller tables by projection. There are basically two goals of normalization as follows:

1. Eliminate redundant data (for example, storing the same data in more than one table)

104) Serious ---- involves decomposing and resembling the data:

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

Data cleansing is much more than simply updating a record with good data. Serious data cleansing involves decomposing and reassembling the data. One can break down the

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

A data warehouse is a "subject-oriented, integrated, time varying, non-volatile collection of data that is used primarily in organizational decision making. Typically, the data

106) For large record spaces and large number of records , the run time of the clustering algorithms:

- Prohibitive page 164
- Static

ARSL SHAHZAD MANAGEMENT SCIENCES FACULTY GROUP

- Exponential
- Numerical

computational time. The clustering algorithms have high computational complexity. For large record spaces and large number of records, the run time of the clustering algorithms is prohibitive.

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

Data duplication can result in costly errors, such as:

- False frequency distributions.
- Incorrect aggregates due to double counting.

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

- Completeness page 185
- Uniqueness
- Accessibility
- Consistency

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

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)

Complexity Analysis of BSN Method

- Time Complexity: O(n log n)
  - O (n) for Key Creation
  - O (n log n) for Sorting
  - O (w n) for matching, where $w \leq 2 \leq n$
  - Constants vary a lot

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

- Inheritance page 124
- Aggregation
- Association
- Asset disposal

ARSL SHANAWANA MALIK & MOMILY GROUP

"Slowly" Changing Dimensions.

There can be many examples. For a young customer who is single, then after a while the customer gets married. After sometime there are children, in unfortunate cases the marriage breaks so the customer is separated or the husband dies and the customer becomes a widow. This just does not typically happen overnight but takes a while. Another example is inheritance, consider the example of land. Over a period of time the

111) The ---------- operator proves useful in more complex metrices applicable to the dimensions and accessibility:

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

believability and appropriate amount of data. The max operator proves useful in more complex metrics applicable to the dimensions of timeliness and accessibility.

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

- Bulk insertion page 75
- Single insertion
- Sequential insertion
- No insertion

Typical write operation	Update, insert, delete	Bulk insert, almost no deletion

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

9.3 Other Issues: Adding redundant columns

Other issues include, increase in table size, maintenance and loss of information:

114) ---------- is applicable in Profitability analysis:

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

5.3 Typical Applications of DWH

There are, and there can be many many applications of a data warehouse. It is not possible to discuss all of them. Some representative applications are listed to be discussed as follows:

- Fraud detection.

ARSL SHANI, A & MALIK & MCS FAMILY GROUP

- Profitability analysis.
- Direct mail/database marketing.
- Credit risk prediction.
- Customer retention modeling.
- Yield management.
- Inventory management.

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

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

relatively static pattern of utilization. There is an essentially different pattern of hardware utilization in the data warehouse environment i.e. a binary pattern of utilization, either the hardware is utilized fully or not at all. Calculating a mean utilization for a DWH is not a

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

Time variant: Time variant records are records that are created as of some moment in time. Every record in the data warehouse has some form of time variancy associated with it. In an OLTP system, the contents change with time i.e. updated such as bank account balance or mobile phone balance, but in a warehouse as the data is loaded; the moment usually becomes its time stamp.

117) In which class of aggregates AVERAGE function can be placed:

- Algebraic page 120
- Distributed
- Associative
- Holistic

- Algebraic
  - Compute aggregate from constant-sized summary of subgroup
  - Examples: STDDEV, AVERAGE
  - For AVERAGE, summary data for each group is SUM, COUNT

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)

ARSL SHANI,AINA & MCS FAMILY GROUP

- Emp_Designation

Although the table is shown in 1NF i.e. it contains atomic values, there are no repeating values, there is no aggregation, yet it contains redundant data. For example, information

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

- Intelligence
- Experience
- Knowledge page 11
- Power

questions such as "who", "what", "where", and "when". Knowledge, on the other hand is an application of information and data, and gives an insight by answering the "how"

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

Goals of Horizontal splitting:

There are typically two specific goals for horizontal partitioning: (1) spread rows in a large table across many HW components (disks, controllers, CPUs, etc.) in the environment to facilitate parallel processing, and (2) 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. Of course, what we would like in an ideal

ARSL SHANI,AINA MALIK ICS FAMILY GROUP

CS614

1) One of the possible issues faced by web scopping is that:

- **Web pages may contain junk data**
- Web pages does not contain multiple facts
- Web pages do not contain multiple dimension
- Web pages does not support transformation

2)De-normalization is the process of selectively transforming normalized relation into un-normalized physical records specifications with the aim to:

- Well structure the data
- Well model the data
- **Reduce query processing time**
- None of the given options

3)_____Allows download of “cube” structures to a desktop platform without the need for shared relational or cube server.

- MOLP
- ROLAP
- **DOLAP**
- HOLAP

4) History is excellent predictor of the_____

- Past
- Present
- **Future**
- History

5) The data in the data warehouse is_____

- Volatile
- **Non-volatile**
- Static
- Non-structure

6)Which of the following is an example of Non-Additive Facts?

- Quantity sold
- Total sale in Rs
- **Discount in percentage**
- Count of orders in a store

7)In_____SQL generation is vastly simplified for front-end tools when the data is highly structure.

- MOLAP
- **Star Scheme**
- Hybrid schema
- Object oriented scheme

8) Which of the following is NOT an example of a typical grain?

- Individual Transactions
- Daily aggregrates
- Monthly aggregates
- **Normalized attributes**

9)In a fully normalized form:

- **Too many joins are required**

- Relationships lose their significance
- No joins are required
- Data integrity becomes an issue

10) In context Change Data Capture(CDC),some times a _____object can be sed to store the recently modified data

- Buffer table
- **Changing table**
- Checkmark table
- Change control table

11) 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**
- Relation R is in 2nd Normal Form
- Relation R is I 3rd Normal Form
- Relation R is I 3rd Normal Form but not in 2nd Normal Form

12) “Sometimes during data collection complete entities are missed”. This statement is an example of:

- **Missing tuple**
- Missing attribute
- Missing aggregation
- Semantically dirty data

13) An OLTP system is always good at_____

- **Evolving data**

- Keeping Static data
- Tracking past data
- Maintaining historic data

14) The trade-off of denormalization is/are:

- Storage
- Performance
- Ease-of-use
- **All of the given option**

15) Online Extraction is a kind of_____ data extraction.

- **Physical**
- Logical
- Dimension
- Multivalued

16) Which kind of relationships is captured by Fact less fact table?

- **Many-to-many**
- One-to-many
- One-to-one
- None of the given options

17) Suppose the size of the attribute “Computerize National Identity Card (CNIC) no” is changed in NADRA database .This transformation refers to:

- **Format revision**
- Field spilitting
- Field decoding

- Calculation of derived value

18) Robotic libraries arec needed for_____

- Cubes
- Data marts
- **Data warehouse**
- Aggregation

19) De-normailization affects:

- **Database size and qery performance**
- Database Usability and query reliability
- Database availability and query success
- None of the given

20) Which people crities Dimensional Mmodeling(DM) as being a data mart oriented approach?

- Those that consider ER model as Data marts
- **Those that consider business process as Data marts**

21) _____faciltates a mobile computing paradigm.

- **DOLAP**
- HOLAP
- ROLAP
- MOLAP

22) Change Data Capture(CDC) can be challenging task because:

- Aggregates don’t change in real time

- **Identifying thr recently modified data may be difficult**

23) Table collapsing technique is applied in case of

- One-to-many relation
- **One-to-one relation or mant-to-many relation**

24) The data has to be checked, cleaned and transformed into a_____ format to allow easy and fast access.

- **Unified**
- Predicted
- Qualified
- Proactive

25) _____ incorporates the concept of product quality, process control, quality assurance, and quality improvement.

- **Total Quality Management**
- Intrinsic Data Quality Management
- Realistic Data Quality Management
- Strong Data Quality Management

26) The extent to which data is in appropriate languages, symbols and units, and the definitions are clear is known as _____.

- **Interpretability**
- Uniqueness
- Accessibility
- Consistency

27)The degree to which values are present in the attributes that require them is known as _____.

- **Completeness**
- Uniqueness
- Accessibility
- Consistency

28) The _____ dimension represents data correctness.

- **Free-of-error**
- Completeness
- Consistency
- Correctness

29) In B-tree index, the lowest level index blocks are called leaf blocks, and these blocks contain:

- NULL value to make the leaf terminal node
- **Every indexed data value and a corresponding ROWID**
- Every indexed data value and pointer to next level block
- Every indexed data value and pointer to root block

30) Data is the _____ on which a Data Warehouse (DWH) runs.

- **Fuel page**
- Element
- Component
- Entity

31) In context of data parallelism to get a speed-up of N with N partitions, it must be ensured that:

- There are enough computing resources
- Query-coordinator is very fast as compared to query servers
- Work done in each partition almost same
- **All of the given options**

32) Which of the following is not an activity of Data Quality Analysis Project?

- "Define"
- "Measure"
- "Analyze"
- **"Compression"**

33) Which of the following is not a Data Quality Validation Technique?

- Referential Integrity
- Using Data Quality Rules
- Data Histograming
- **Indexes**

34) One of the preconditions to decide about operations to be parallelized is that

- **Operation can be implemented independent of each other**
- Output of one operation becomes input of other
- Operations share same memory location
- Operations share same namespace

35)_____do not (typically) keep the index values in sorted order

- Dense index
- Sparse index
- B-Tree Index
- **Hash Based index**

36) Parallelism can be exploited, if there is :

- Symmetric multi processors (SMP)

- Sufficient I/O bandwidth
- Underutilized or intermittently used CPUs
- **All of the given options**

37)Which of the following is NOT one of the parallel hardware architectures?

- Symmetric Multi-Processing
- Massively Parallel Processing
- Non-uniform Memory Access
- **Shared Memory**

38)Two interesting examples of quality dimensions that can make use of the min operator are _____.

- **Believability and appropriate amount of data**
- Believability and Consistency
- Believability and Redundancy
- Reliability and appropriate amount of data

39) As the number of processors increase the speedup should also increase. Thus we should have linear speedup. Which of the following is NOT one of the barriers to achieve this linear speed-up?

- Amdahl Law
- Startup
- **No Interference**
- Skew

40) In _____index, the ith bit is set to "1" if the ith row of the base table has the value for the indexed column

- Inverted index
- **Bitmap index**
- Cluster index
- Join index

41)_____ lists each term in the collection only once and then shows a list of all the documents that contain the given term.

- **Inverted index page**
- Bitmap index
- Cluster index
- Join index

42) The exact formula for Speed-up is:

- **(Time on Serial Processor) / (Time on parallel processors)**
- (Time on Serial Processor) * (Time on parallel processors)
- (Time on Serial Processor) + (Time on parallel processors)
- (Time on Serial Processor) - (Time on parallel processors)

43) _____is the degree to which data accurately reflects the real world object that the data represents

- **Intrinsic data quality**
- Realistic data quality
- Strong data quality
- Weak data quality

44) Assume a company with a multi-million row customer table i.e. n rows. Checking for Referential Integrity (RI), using a smart technique with some kind of a tree data structure would require _____ time.

- **O(log n)**
- O(n)
- O(1)
- None of the given

45) Which of the following is NOT one of the variants of Nested-loop join?

- Naive nested-loop join
- Indexed nested-loop join
- Temporary index nested-loop join
- **Binary index nested-loop join**

46). “More resources means proportionally less time for given amount of data” that statement refers to

- Scale-Up
- **Speed-Up**
- Size-Up
- Over-Utilized system

47)The optimizer uses a hash join to join two tables if they are joined using an equijoin and

- outer table has less number of rows
- inner table has less number of rows
- cardinality of table is equal
- **large amount of data needs to be joined**

48) “If resources increase in proposition to increase in data size, time is constant”. The statement refers to

- **Scale-up**
- Speed-up
- Size-up
- Over-utilized system

49) If a product meets formally defined “requirement specifications”, yet fails to be a quality product form the customer’s perspective, this means the requirements were _____.

- **Defective**
- Unclear
- Unrefined

- Undefined

50)_____is the extent to which data is regarded as true and credible.

- **Believability**
- Completeness
- Accessibility
- consistency

51) Which is not a/an data quality validation technique?

- **Consistency integrity**
- Referential integrity
- Attribute domain
- Using data quality rules

52) Which of the following is not an "Orr's law of data quality"?

- Data that Is not used is cannot be correct
- Data quality is 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**

53)_____ is known as state of being only one of its kind or being without an equal or parallel.

- Completeness
- **Uniqueness**
- Accessibility
- Consistency

54) Which is not a/an characteristics of data quality?

- **Reliability**
- Uniqueness

- Accessibility
- Consistency

55) If every key in the data file is represented in the index file then it is called

- **Dense Index**
- Sparse Index
- Inverted Index
- A Multi level Sparse Index

56)One of the main reasons for the failure of DWH deployment is ____

- **Data quality**
- Data integrity
- Data duplication
- Data anomaly

57) The _____ operator is conservative in that it assigns to the dimension an aggregate value no higher than the value of its weakest data quality indicator.

- **Min**
- Max
- Min and Max
- None of given

58) ____ is making all efforts to increase effectiveness in meeting and deficiency in meeting except customer expectations

- Quality assurance
- **Quality improvement**

Quality maintenance
Quality establishment

59) Most DWH implantations today do not use ____ enforced by the database, but as TQM method improved overall data quality and database optimizers.

- Consistency integrity

- **Referential integrity**
- Attribute domain
- Using data quality rules

60) If a task takes "T" time units to execute on a single data item, then execution of the Task on "N" data items will take _____time units.

- **N*T**
- N/T
- N+T
- N-T

61) An optimized structure which is built primarily for retrieval, with update being only a secondary consideration is

- OLTP
- OLAP
- DSS
- **Inverted Index**

62)_____ refers to "Parallel execution of single data operation across multiple partitions of data"

- Hardware parallelism
- Software parallelism
- **Data parallelism**
- Operational parallelism

63)_____ in a database or data warehouse has no actual value, it only has potential value.

- **Data**
- Entity
- Flat tables

- Data marts

64) Which of the following tasks can NOT be parallelized?

- Large table scans and joins
- Creation of large indexes
- Partitioned index scans
- **None of the given options**

65)A join is identified by multiple tables in the _____ clause

- **FROM**
- SELECT
- GROUP BY
- SORT BY

66). _____ index stores first value in each block in the sequential file and a pointer to the block.

- Dense
- **Sparse**
- B-Tree
- Hash

67) _____ is a/an measure of how current or up to date the data is

- **Timeliness**
- Completeness
- Accessibility
- Consistency

68) In context of data parallelism, the work done by query processor should be:

- **Almost zero**
- Maximum
- Pipelined
- Filtered across partitions

69)In context of joining tables, the join condition is specified in _____clause

- FORM
- SELECT
- **WHERE**
- GROUP BY

70)A _____ index, if fits in the memory, costs only one disk I/O access to locate a record given a key.

- **Dense**
- Sparse
- B-Tree
- Hash

71) _____ index uses even less space than _____ index, but the block has to be searched, even for unsuccessful searches.

- Dense, sparse
- **Sparse, dens**
- Dense, inverted
- Sparse, inverted

72)_____ is the degree of utility and value the data has to support the enterprise processes that enable accomplishing enterprise objectives.

- Intrinsic Data Quality

- **Realistic Data Quality**
- Strong Data Quality
- Weak Data Quality

73) _____ is a system of activities that assures conformance of product to pre-established requirements.

- **Quality assurance**
- Quality improvement
- Quality maintenance
- Quality establishment

74) In context of nested-loop join actual number of matching rows returned as a result of the join would be _____ of the order of tables

- Dependent
- **Independent**
- Superset
- Subset

75) In context of bitmap index, the length of the bit vector is:

- The possible number of domain values in corresponding field (column)
- **The number of records in the base table**
- The possible number of bitmap tables formed for corresponding field (column)
- None of the given options

76)The _____ operator proves useful in more complex metrics applicable to the dimensions of timeliness and accessibility.

- **Max page**
- Min
- Min and Max
- None of given

77) In nested-loop join case, if there are ‘M’ rows in outer table and ‘N’ rows in inner table, time complexity is

- (M log N)
- O (log MN)
- **O (MN)**
- O (M + N)