



Virtual University

CS614- DATA
WEARHOUSING
(SOLVED MCQs)
FROM MIDTERM PAPERS
LECTURE (1-22)



Al Junaid Tech®

ajtechinstitute24@gmail.com

For More Visit: vujunaid.com

JUNAID MALIK (0304-1659294)



Al Junaid Tech®



پروجیکٹ اور انٹرنشپ
ساتھ ساتھ

Daily Classes & Viva
Preparation

let's keep **Learning** and **growing**
together!



Our Expertise

- Python (Django & Flask)
- Php (Laraval)
- MySql

Projects Phase

- SRS + DD + Prototype
- Final Deliverable

LMS
Handling

+92 304 1659294

ajtechinstitute24@gmail.com

AL-JÛÑÂÎD TÊCH

JOIN OUR GROUP FOR MORE UPDATE

[Click Button](#)

CS619 & CS519

ÂL-JÛÑÂÎD

1. Online Extraction is a kind of-----data extraction.
 - Logical
 - Dimensional
 - **Physical page 132**
 - Multi valued
2. The ----- saw the advent of disk storage, or DASD(direct Access Storage Device) :
 - 1960s
 - **1970s page 13**
 - 1950s
 - 1990s
3. In context of data warehouse, normally it becomes difficult to extract data from different sources because these sources are normally.
 - **Heterogeneous page 140**
 - Homogeneous
 - Centralized
 - Baseline
4. 5Which of the following is not a task of Data Transformation?
 - Conversion
 - Summarization
 - Enrichment
 - **Full Data Refresh page 135**
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**

AL-JÛÑÂÎD TÊCH

6. Flat files are one of the prevalent structures used in ----- data extraction:
- Online
 - **Offline** page 134
 - Incremental
 - Full
7. Which of the following is NOT one of the advantages of changed data capture (CDC) technique?
- Flat files are not required
 - **Limited query interface is required for data extraction** page 152
 - No incremental on-line I/O required for log tape
 - Extraction of changed data occurs immediately
8. The most common range partitioning is on
- Color
 - **Date** page 66
 - PhoneNo
 - Name
9. A relation is said to be in first normal form(1NF), if it does not contain _____
- Single value column
 - **Multi-valued column** page 43
 - Derived column
 - Composite column
10. In a fully normalized database, too many _____ are required
- Values
 - **Joins** page 49
 - Queries
 - Conditions
11. In the data warehouse, data is collection from ----- sources:
- Homogeneous
 - **Heterogeneous** page 21
 - External
 - Internal
12. De-normalization is more like a “controlled crash” with the aim to ----- without loss of information:
- Check
 - Balance
 - Decrease

AL-JÛÑÂÎD TÊCH

- **Enhance** page 49
13. 1----- is making all efforts to increase effectiveness and efficiency in meeting and accepted customer expectation:
- Quality assurance
 - **Quality improvement** page 183
 - Quality maintenance
 - Quality Establishment
14. ----- is the application of intelligence and experience to get common goals.
- **Wisdom** page 11
 - Education
 - Power
 - Information
15. 1In the data transformation, ----- is the rearrangement and simplification of individual
- Aggregation
 - **Enrichment** page 136
 - Splitting joining
 - Conversion
16. Grain of a fact table means :
- **The meaning of one fact table row** page 109
 - The meaning of one dimensional table row
 - Summary of aggregates in all fact tables
 - Summary of aggregates in all dimension tables
17. Normalization ----- :
- **Reduces redundancy** page 41
 - Increases redundancy
 - Reduces joins
 - Reduces tables
18. 1Which of the following is not an example of a typical grain :
- Individual transaction
 - Daily aggregates
 - Monthly aggregates
 - **Normalized attributes** page 111
19. Multi-dimensional databases(MDDs) typically use ----- formats to store pre-summarized cube structures:

AL-JÛÑÂÎD TÊCH

- SQL
 - **Proprietary file** page 79
 - Object oriented
 - Non-proprietary file
20. ----- provides a combination of “relational databases access” and “cube” data structures within a single framework:
- **HOLAP** page 78
 - DOLAP
 - MOLAP
 - ROLAP
21. Data Warehouse provides the best support for analysis while OLAP carries out the ----- task:
- Mandatory
 - Whole
 - **Analysis** page 69
 - Prediction
22. ----- involves splitting a table by columns so that a group of columns is placed into the new table and the remaining columns are placed in another new table:
- **Vertical splitting** page 56
 - Horizontal splitting
 - Adding redundant column
 - None of the given option
23. OLAP implementations are highly/completely ----- :
- Normalized
 - **Demoralized** page 69
 - Predictive
 - Additive
24. If each cell of Relation R contains a single value (no repeating values) then it is confirmed that :
- **Relation R is in 1st Normal Form** page 43
 - Relation R is in 2nd Normal Form
 - Relation R is in 3rd Normal Form
 - Relation R is in 3rd Normal Form but not in 2nd Normal Form
25. Which kind of relationships is captured by Fact less fact table:
- **Many- to- Many** page 121
 - One-to-many

AL-JÛÑÂÎD TÊCH

- One-to-one
 - None of the given option
26. Which of the following is not an example of dimension:
- Product
 - Date
 - Region
 - **Sales volume** page 78
27. Which people criticize Dimensional Modeling (DM) as being a data mart oriented approach?
- Those that consider ER models as Data marts
 - **Those that consider Business processes as Data marts** page 110
 - Those that consider Data marts as Data warehouse
 - Those that consider dimensional model
 - Those that consider dimensional modeling as de-normalization approach
28. In a fully normalized form:
- **To many joins are required** page 49
 - Relationships lose their significance
 - No joins are required
 - Data integrity becomes an issue
29. Which of the following is an example of Non-Additive Facts:
- Quality sold
 - Total sale in Rs.
 - **Discount in percentage** page 119
 - Count of orders in a store
30. Which of the following is not a CUBE operation?
- **ANSI SQL** page 81
 - Roll Up
 - Drill Down
 - Pivoting
31. ----- allows download of “cube” structures to a desktop platform without the need for shared or cube server:
- MPLAP
 - ROLAP
 - **DOLAP** page 78

AL-JÛÑÂÎD TÊCH

- HOLAP
32. ROLAP provides access to information via a relational database using:
- **ANSI standard SQL** page 78
 - Proprietary file format
 - Comma Separated Values
 - All of the given option
33. ----- is usually deployed when expression can be used to group data together in such a way that access can be targeted to a small set of partitions:
- Expression elimination
 - **Expression partitioning** page 67
 - Expression indexing
 - None of the given option
34. Taken jointly, the extract programs or naturally evolving systems formed a spider web, also known as
- Distributed Systems Architecture
 - **Legacy System Architecture** page 14
 - Online System Architecture
 - Intranet System Architecture
35. The data has to be checked , cleaned and transformed into a ----- format to allow easy and fast access
- **Unified** page 20
 - Predicated
 - Qualified
 - Proactive
36. Suppose in a system A, the values of “PhoneNo” attribute were stored in “countrycode-phone-extension” format, however after transformation into data warehouse the separate columns were used for “countrycode”,”phone” and “extension”. The above scenario is an example of :
- One-to-one scalar transformation
 - **One-to-many element transformation** page 144+conceptual
 - Many-to-one element transformation
 - Many-to-many element transformation
37. In decision support system ease of use is achieved by:
- Normalization
 - **Denormalization** page no 49

AL-JÛÑÂÎD TÊCH

- Drill up
 - Drill down
38. Which of the following is one of the methods to simplify an ER model?
- Normalization
 - **Denormalization** page no 103
 - HOLAP
 - Hybrid schema
39. In ETL process data transformation includes -----
- **Data cleansing** page 129
 - Data aggregation
 - Behavior checking
 - Pattern recognition
40. Non-uniform use of abbreviations, units, and values refers to:
- **Syntactically dirty data** page 160
 - Semantically dirty data
 - Coverage anomaly
 - Extraction issue
41. Suppose the size of the attribute “Computerized National Card (CNIC) no. is changed in NADRA database. This transformation refers to:
- **Format revision** page 153
 - Field splitting
 - Field decoding
 - Calculation of derived value
42. The divide and conquer cube partitioning approach helps alleviate the ----- limitations of MOLAP implementation:
- Flexibility
 - Maintainability
 - Security
 - **Scalability** page 85
43. identify the TRUE statement:
- DM is inherently dimensional in nature
 - DM comprises of a single central fact table
 - DM comprises of a set of dimensional tables
 - **All of the given option** Page 103
44. ----- can be used when some columns are rarely accessed rather than other columns or when the table has wide rows or header or both:

AL-JÛÑÂÎD TÊCH

- Horizontal splitting
 - Pre-joining
 - **Vertical splitting** page 56
 - Derived attributes
45. Which of the following is an example of derived attributes?
- **Age** page 61
 - Size
 - Color
 - Length
46. The online high performance transaction processing was evolved in -----:
- 1980
 - **1975** page 12
 - 1977
 - 1965
47. Cube is a logical entity containing values of a certain aggregation level at an intersection of a combination of ----- :
- Facts
 - **Dimension** page 88
 - Summary tables
 - Primary and foreign key
48. Which of the following is TRUE regarding Entity relationship modeling?
- It does not really model business, but models the micro relationships among data elements.
 - ER modeling does not have “business rules,” it has “data rules
 - ER modeling helps retrieval of individual records having certain critical identifiers.
 - **All of the given option** page 102
49. -----Facilitates a mobile computing paradiagramn:
- HOLAP
 - **DOLAP** page78
 - ROLAP
 - MOLAP
50. The main reason(s)for the increase in cube size may be:
- Increase in the number of dimensions
 - Increase in the cardinality of the dimensions

AL-JÛÑÂÎD TÊCH

- Increase in the amount of detail data
- **All of the given options** page 87

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

- Linear
- Quadratic
- **Exponential** page 15
- Logarithmic

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

- Volatile
- **Non-volatile** page 69
- Static
- Non-structured

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

- **Dimensional model** page 102
- Entity relationship model
- Object oriented model
- Structured model

54. ----- 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

55. Which of the de-normalization technique squeezes master table into detail?

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

56. De-normalization can help:

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

AL-JÛÑÂÎD TÊCH

57. 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

58. Increasing level of normalization --- number of
----- tables:

- **Increases** page 51
- Decreases
- Does not effect
- None of the given option

59. Which of the following is not a Data Quality Validation Technique:

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

60. 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

61. 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

62. 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

63. History is excellent predictor of the -----:

- Past

AL-JÛÑÂÎD TÊCH

- Present
- **Future** page 15
- History

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

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

65. ----- gives total view of an organization:

- OLAP
- OLTP
- **Data Warehouse** page 16
- Database

66. 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

67. Enrichment is one of the basic tasks in data ----- :

- Extraction
- **Transformation** page 138
- Loading
- Summarization

68. Which of the following is not a technique of De-normalization?

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

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

- **Sales Amount** page 119
- Average
- Discount
- Ratios

AL-JÛÑÂÎD TÊCH

70. Robotic libraries are needed for -----:

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

71. Normally ROLAP is implemented using -----

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

72. 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

73. In --

---- **ested loop join of quadratic time complexity does not hurt the**
---- **performance**

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

74. 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

75. 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
- $O(n)$

AL-JÛÑÂÎD TÊCH

- $O(n^2)$

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

- **Smaller** page 84

- Larger
- Partitioned
- Merged

77. 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

78. The trade-offs of de-normalization is/are:

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

79. 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

80. “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

-----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

AL-JÛÑÂÎD TÊCH

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

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

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

- Age
- CGPA
- Annual Salary

83. 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

84. MDX by Microsoft is an example of -----:

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

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

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

86. 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

87. OLAP is:

AL-JÛÑÂÎD TÊCH

- **Analytical processing** page 69

- Transaction processing
- Additive processing
- Active processing

88. 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

89. Which of the following is\are example of dimension:

- **Product** page 79

- Region
- Data
- None of the given

90. An OLTP system is always good at -----:

- **Evolving data** page 122

- Keeping static data
- Tracking past data
- Maintaining historic data

91. 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

92. 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

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

AL-JÛÑÂÎD TÊCH

- MOLAP
 - **Star Schema** page 107
 - Hybrid schema
 - Object oriented schema
94. Dirty data means:
- Data cannot be aggregated
 - Data contains non-additive facts
 - Data does not fulfill dimensional modeling rules
 - **Data does not conform to proper domain definitions** page 158
95. 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
96. “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
97. 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
98. Which of the following is an example of dimension?
- Product
 - Region
 - Date
 - **All of the given option** page 78
99. Data warehouse stores -----:
- Operational data
 - **Historical data** page 24
 - Meta data

AL-JÛÑÂÎD TÊCH

- Log files data
100. 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
101. The main goal of normalization is to eliminate -----:
- **Data redundancy** page 41
 - Data sharing
 - Data security
 - Data consistency
102. Serious ---- involves decomposing and reassembling the data:
- **Data cleansing** page 168
 - Data transformation
 - Data loading
 - Data extraction
103. 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
104. For large record spaces and large number of records , the run time of the clustering algorithms:
- **Prohibitive** page 164
 - Static
 - Exponential
 - Numerical
105. ----- 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
106. The degree to which values are present in the attributes that require them is known as -
-----:

AL-JÛÑÂÎD TÊCH

- **Completeness** page 185
 - Uniqueness
 - Accessibility
 - Consistency
107. Time complexity of Key Creation process in basic Sorted Neighborhood (BSN) Method is -----:
- $O(n \log n)$
 - $O(\log n)$
 - **$O(n)$** page 171
 - $O(2n)$
108. Which of the following is an example of slowly changing dimensions?
- **Inheritance** page 124
 - Aggregation
 - Association
 - Asset disposal
109. The ----- operator proves useful in more complex metrics applicable to the dimensions and accessibility:
- **Max** page 188
 - Min
 - Max and Min
 - None of the given
110. In OLAP, the typical write operation is ----- :
- **Bulk insertion** page 75
 - Single insertion
 - Sequential insertion
 - No insertion
111. The issue(s) of “ Adding redundant column” includes(s):
- Increase in table size
 - Maintenance
 - Loss of information
 - **All of the given option** page 65
112. ----- is applicable in Profitability analysis:
- OLTP
 - **Data warehouse** page 36,37

AL-JÛÑÂÎD TÊCH

- Information System(IS)
- Management Information System(MIS)

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

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

114. 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

115. In which class of aggregates AVERAGE function can be placed:

- **Algebraic page 120**
- Distributed
- Associative
- Holistic

116. Considered the following Employee table and identify the column which causes that the table is not in first normal form(1NF): (Emp_ID, Emp_Name ,Emp_skills, Emp_Designation)

- Emp_ID
- Emp_Name
- **Emp_skills page 43(conceptual)**
- Emp_Designation

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

- Intelligence
- Experience
- **Knowledge page 11**
- Power

118. ----- 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

AL-JÛÑÂÎD TÊCH

- Collapsing table technique
 - **Horizontal splitting technique page 56**
 - Vertical splitting technique
119. -----should not be present in a relation, so that it would be in second normal form (2NF).
- **Partial dependency page 44 (conceptual)**
 - Full functional dependency
 - Multivalued dependency
 - Transitive dependency
120. 1Records referring to the same entity are represented in different formulas in the different data sets or are represented erroneously. Thus duplicate records will appear in the merged database. This problem is known as-----.
- **Merge/purge problem page 168**
 - Duplication problem
 - Redundant duplication problem
 - Redundant problem
121. The data perspective in OLTP system is operational, while that in data warehouse is:
- Fully normalized
 - Fully de-normalized
 - Fully summarized
 - **Historical and detailed page 30**
122. 1Simple scalar transformation is a-----mapping from one set of values to another set of values using straightforward rules.
- **One-to-one page 144**
 - One-to-many
 - Many-to-many
 - Many-to-one
123. -----can be created in operational systems to keep tracks of recently updated records.
- **Triggers page 150**
 - Timestamps
 - Partitioning
 - ELT
124. 1Development of data warehouse is hard because data sources are usually-----

AL-JÛÑÂÎD TÊCH

- Structured and homogeneous
 - **Unstructured and heterogeneous page 31**
125. In a decision support environment, the decision maker is interested in -----.
- Structured and heterogeneous
 - Unstructured and homogeneous
 - Only limited organizational data
 - **Big picture of organizational data page 21**
 - Only sale related data
 - Only customer related data
126. Information can answer question like “what”, “who” and “when” while knowledge can answer question like-----.
- Why
 - Where
 - Which
 - **How page 11**
127. OLTP implementations are fully-----.
- **Normalized page 69**
 - Denormalized
 - Predictive
 - Additive
128. Which logical data extraction has significant performance impacts on the data warehouse server?
- **Incremental Extraction page 133**
 - Online Extraction
 - Offline Extraction
 - Legacy Vs OLTP
129. Consider the following Student table and identify the column which causes that the table is not in first normal form(1NF).
Student(Std_ID, Std_Name ,Std_CGPA ,Std_Hobbies)
- Std_ID
 - Std_Name
 - Std_CGPA
 - **Std_Hobbies page 43(Conceptual)**
130. Analytical processing uses -----
- **Multi-level aggregates page 74**

AL-JÛÑÂÎD TÊCH

- Record level aggregates
 - Table level aggregates
 - All of the given options
131. Which is not a class of anomalies in following?
- **Dirty anomalies** page 160
 - Syntactically dirty data
 - Semantically dirty data
 - Coverage anomalies
132. _____ is a system of activities that assures conformance of product to pre-established requirements.
- **Quality assurance** page 183
 - Quality improvement
 - Quality Maintenance
 - Quality Establishment
133. Two interesting examples of quality dimensions that can make use of min operator are _____
- **Believability and appropriate amount of data** page 188
 - Believability and consistency
 - Believability and Redundancy
 - Reliability and appropriate amount of data
134. _____ in database or data warehouse has no actual value; it only has potential value.
- **Data** page 181
 - Entity
 - Flat tables
 - Data marts
135. In OLTP environment the selectivity is _____ and _____ in data warehouse environment.
- **High, Low** page 22
 - Low, High
 - High, Fixed
 - Fixed, Low
136. Which is not a/an characteristics of data quality?
- **Reliability** page 186
 - Uniqueness

AL-JÛÑÂÎD TÊCH

- Accessibility
 - Consistency
137. 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 -----.
- **Defective page 180**
 - Unclear
 - Unrefined
 - Undefined
138. The relation R will be in 3rd Normal Form if:
- It is in 2NF each cell contains single value
 - **It is in 2NF and every non-key column is non-key transitively dependent upon its primary key. Page 46**
 - It is in 1NF and each non key attribute is dependent upon a single column of composite primary key.
 - It is in 2NF and each non key attribute is dependent upon other non-key attribute.
139. Decision support system queries deal with number of columns -----
- Having numeric values
 - In a single table
 - In a single view
 - **Spanning across multiple tables page 21**
140. Normalization is used to reduce:
- **Reduces redundancy page 41**
 - Increases redundancy
 - Reduces joins
 - Reduces tables
141. The end user of data ware house are-----.
- Programmers
 - Database developers
 - Data entry operator
 - **Business executives page 18 + 19**
142. Which one are the characteristics of data warehouse queries?
- Use primary key
 - High selectivity

AL-JÛÑÂÎD TÊCH

- **Use multiple tables** page 30
 - Very low performance
143. 1Referential Integrity (RI) using a naive approach would take ----- time.
- **O(n)** page 160
 - O(1)
 - O(log n)
 - None of the given
144. Web scrapping is a process of applying ----- techniques to the web
- **Screen scrapping** page 146
 - Data scrapping
 - Text scrapping
 - Meta scrapping
145. Which is not an issue of ROLAP in the following?
- **Standard hierarchy of dimensions** page 92
 - Non-standard conventions
 - Maintenance
 - Aggregation
146. One of the fundamental purpose of de-normalization is to ----- a number of physical tables which ultimately reduce the number of joins to answer a query.
- Delete
 - **Reduce** page 50
 - Increase
 - Decrease
147. ----- is not the characteristic of data warehouse.
- Time variant
 - Subject-oriented
 - Integrated
 - **Volatile** page 69
148. Which is not a/an step of data cleansing procedure?
- **Aggregation** page 168
 - Elementizing
 - Standardizing
 - Verifying

AL-JÛÑÂÎD TÊCH

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

on identifying attribute(s), if you are very, very, very lucky.
- **Equi-join** page 169
 - Inner join
 - Outer join
 - Fuller join
150. Ad-hoc access of data warehouse means:
- That have predefined database access pattern
 - **That does not have predefined database access pattern** page 18
 - That could be accessed by any user
 - That could not be accessed by any user
151. In OLTP environment, the size of tables is relatively-----
- Large
 - Fixed
 - Moderate
 - **Small** page 22
152. 1----- is a/an measure of how current or up to date the data is:
- **Timeliness** page 185
 - Completeness
 - Accessibility
 - Consistency
153. 1The process of converting entity relationship model in to dimensional model of -

steps:
- Two
 - Three
 - **Four** page 109
 - Five
154. 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).
- **Pattern** page 164
 - Cluster
 - Entity
 - Attribute

AL-JÛÑÂÎD TÊCH

155. ----- is known as state of being only one of its kind or being without an equal or parallel.

- Completeness
- **Uniqueness** page 185
- Accessibility
- Consistency

156. Which of the following is not an example of fact?

- **Account no** page 74
- Sales quantity
- Per unit sales amount
- Sales amount

157. 1-----is the degree to which data accurately reflects the real world object that the data represents.

- **Intrinsic data quality** page 181
- Realistic data quality
- Strong data quality
- Weak data quality

158. 1Which one among the following data warehouse stores data containing long period?

- Telecommunication data warehouse
- Financial data warehouse
- Human resource data warehouse
- **Insurance data warehouse** page 36

159. A _____ dimension is a collection of random transactional codes, flags and/text attributes that are unrelated to any particular dimension. The _____ dimension is simply a structure that provides a convenient place to store the _____ attributes.

- **Junk**
- Time
- Parallel
- None of these

160. 1Data Warehouse is about taking / collecting data from different _____ sources.

- Harmonized

AL-JÛÑÂÎD TÊCH

- Identical
 - **Homogeneous**
 - Heterogeneous
161. Taken jointly, the extract programs or naturally evolving systems formed a spider web, also known as
- Distributed Systems Architecture
 - **Legacy Systems Architecture**
 - Online Systems Architecture
 - Intranet Systems Architecture
162. It is observed that every year the amount of data recorded in an organization
- **Doubles**
 - Triples
 - Quartiles
163. The users of data warehouse are knowledge workers in other words they are _____ in the organization.
- DWH Analyst
 - **Decision maker**
 - Database Administrator
 - Manager
164. Node of a B-Tree is stored in memory block and traversing a B-Tree involves _____ page faults.
- $O(n \lg n)$
 - **$O(\log n)$ { $O(\log n)$ it's the real answer }**
 - $O(n)$
 - $O(n^2)$
165. In _____ system, the contents change with time.
- **OLTP**
 - ATM
 - DSS
 - OLAP
166. The growth of master files and magnetic tapes exploded around the mid-_____.
- 1950s.
 - **1960s.**
 - 1970s.
 - 1980s.
167. Relational databases allow you to navigate the data in _____ that is appropriate using the primary, foreign key structure within the data model
- Only One Direction
 - **Any Direction**
 - Two Direction
 - None of these

AL-JÛÑÂÎD TÊCH

168. Naturally Evolving architecture occurred when an organization had a _____ approach to handling the whole process of hardware and software architecture.

- **Relaxed**

- Good
- Not Relaxed
- None

169. _____ gives total view of an organization

- OLAP
- OLTP

- **Data Warehouse**

- Database

170. Suppose the amount of data recorded in an organization is doubled every year. This increase is

- Linear
- Quadratic

- **Exponential**

- Logarithmic

171. _____ is the application and experience to get common goals.

- **Wisdom**

- Power
- Information
- Education

172. Normally ROLAP is implemented using _____.

- **Star Schema**

- Pre-defined aggregates
- Hybrid Schema
- All of the given options

173. Identify data warehouse query from the following.

- Number of students studying CS614 course
- **Factors which can effect students' performance**

- List of students belong to Lahore city
- Total number of students that have 3.5 CGPA

174. In _____ SQL generation is vastly simplified for front-end tools when the data is highly structured.

- Hybrid Schema

- **Star Schema**

- MOLAP
- Object oriented schema

AL-JÛÑÂÎD TÊCH

175. _____ models the Marco relationships among data elements with an overall deterministic strategy.

➤ **Dimensional model**

- Structured model
- Object oriented model
- Entity relationship model

176. The _____ task is typically performed after most other transformation and cleaning steps have taken place.

- Data extraction

➤ **Data duplicate elimination**

- Data transformation
- Data loading

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

- Completeness
- Consistency
- Accessibility

➤ **Uniqueness**

178. Relational modeling techniques are used to develop _____ systems.

➤ **OLTP**

- MOLAP
- OLAP
- ROLAP

179. Which of the following is not CUBE operation?

➤ **ANSI SQL**

- Drill Down
- Pivoting
- Roll UP

180. The hybrid OLAP(HOLAP) solution is a mix of _____ architecture that supports queries against summary and transaction in an integrated.

- ROLAP
- OLTP

➤ **MOLAP and ROLAP**

- MOLAP

181. The response time of OLTP queries are in milliseconds to seconds. While that of data warehouse queries are in:

AL-JÛÑÂÎD TÊCH

- Days of weeks
- Hours to days
- Milliseconds to microseconds

➤ **Minutes to hours**

182. _____ allows download of “cube” structures to a desktop platform without the need for shared relational or cube server.

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

183. In case of multiple sources for the same data element. We need to prioritize the source systems per element bases, the process is called:

- **Ranking**
- Measurement selection
- Element selection
- Prioritization

184. OLAP is:

- Implementation technique
- **Framework**
- None of the given options
- Physical database design

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

- Summarization
- Enrichment
- Conversion
- **Full data refresh**

186. Simple scalar transformation is a _____ mapping from one set of values to another set of values using straightforward rules.

- Many-to-one
- Many-to-many
- **One-to-one**
- One-to-many

187. Which of the following is not a data quality validation technique?

- **Indexes**
- Using data quality rules
- Referential integrity

AL-JÛÑÂÎD TÊCH

- Data Histograming

188. _____ are created out from the data warehouse to service different needs of departments in the organization.

- Databases

➤ **Data marts**

- MIS

- Tables

189. An OLTP system is always good at _____.

- Maintaining historical data

- Keeping static data

- Tracking past data

➤ **Evolving data**

190. _____ incorporates the concept of product quality, process control, quality improvement.

- Intrinsic data quality management

- Realistic Data Quality Management

➤ **Total Quality Management**

- Strong Data Quality Management

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

- Missing aggregate

➤ **Missing tuple**

- Semantically dirty data

- Missing attribute

192. The last step of Software Development Life Cycle (SDLC) is implementation. While that of data warehouse is:

- Integration

➤ **Understanding requirements**

- Testing

- Analysis

193. In which class of aggregates Max function can be places?

- Algebraic

- Associative

- Holistic

➤ **Distributive**

AL-JÛÑÂÎD TÊCH

194. Telecommunication data warehouse is dominated by the _____ volume of data generated at the call level.

- Incomplete
- Semi-incomplete
- **Sheer**
- Partial

195. 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:

- Semantically dirty data
- **Syntactically dirty data**
- Extraction issue
- Coverage anomaly

196. In the data warehouse environment the date is _____.

- **Both subject and time oriented**
- Neither time-oriented nor subject-oriented
- Subject-oriented
- time-oriented

197. **dirty data means that:**

- Data control be aggregated
- Data contain non-additive facts
- Data does not fulfill dimensional modeling rules
- **Data does not conform to proper domain definitions**

198. One of the main reason for the failure of DWH deployments is _____.

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

199. The data in the data warehouse is _____.

- Non-structured
- Volatile
- Static
- **Non-volatile**

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

- Daily aggregates
- Monthly aggregates
- Individual transaction

AL-JÛÑÂÎD TÊCH

➤ Normalized attribute

201. Data warehouse stores ____

- Operational data
- Historical data
- Meta data
- Log files data

202. The ____ dimension represents data correctness

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

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

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

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

- Quantity sold
- Total Sale in Rs.
- Discount Percentage
- Count of orders in a store

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

- Many-to-many element transformation
- One-to-one scalar transformation
- One-to-many element transformation
- All of the given

206. Serious ____ involves decomposing and reassembling the data

- Data cleansing
- Data transformation
- Data loading
- Data Extraction

207. ____ 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

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

- Normalization
- Denormalization
- Drill up

AL-JÛÑÂÎD TÊCH

- Drill down
209. 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 tree data structure would require ___ time
- $O(\log n)$
 - $O(n)$
 - $O(1)$
 - None
210. Which of the following is NOT an example of a typical grain?
- Individual Transactions
 - Daily aggregates
 - Monthly aggregates
 - Normalized attributes
211. Most DWH implementations today do not use ___ enforced by the database, but as TQM methods improve overall data quality and database optimizers
- Consistency Integrity
 - Referential Integrity
 - Attribute domain
 - Using Data Quality Rules
212. 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. The above scenario is an example of:
- One-to-one scalar transformation
 - One-to-many element transformation
 - Many-to-one element transformation
 - Many-to-many element transformation
213. Development of data warehouse is hard because data sources are usually ___
- Structured and homogeneous
 - Unstructured and heterogeneous
 - Structured and heterogeneous
 - Unstructured and homogeneous
214. A/an ___ is a collection of random transactional codes, flags and/or text attributes that are unrelated to any particular dimension
- Junk dimension
 - Slowly changing dimension
 - Multi-valued dimension
 - Simple dimensions
215. ROLAP provides access to information via a relational database using
- ANSI standard SQL
 - Proprietary file format
 - Comma Separated Values
 - All of the given

AL-JÛÑÂÎD TÊCH

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

- 6/12

- 7/12
- 1/24
- Twice a week

217. In ___ nested-loop join of quadratic time complexity does not hurt the performance

- Typical OLTP environments

- Data warehouse
- DSS
- OLAP

218. If actual data structure does not conform to documented formats then it is called:

- Syntactically dirty data

- Semantically dirty data
- Coverage anomaly
- Extraction issue

219. Which of the following is not a CUBE operation?

- ANSI SQL

- Roll UP
- Drill down
- pivoting

220. The data has to be checked, cleansed and transformed into a ___ format to allow easy and fast access

- unified

- predicted
- qualified
- proactive

221. Which is not a/an characteristics of data quality?

- Reliability
- Uniqueness
- Accessibility
- Consistency

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

- Interpretability

- Uniqueness
- Accessibility
- Consistency

223. In case of multiple sources for the same data element, we need to prioritize the source systems per element bases, the process is called

- Ranking

AL-JÛÑÂÎD TÊCH

- Prioritization
 - Element Selection
 - Measurement Selection
224. In OLTP environments, the size of tables is relatively ____
- Large
 - Fixed
 - Moderate
 - **Small**
225. Change Data Capture (CDC) can be challenging task because
- Aggregates don't change in real time
 - Transformation of extracted data is difficult
 - **Identifying the recently modified data may be difficult**
 - Source systems may not support extraction of changed aggregates
226. ____ is the extent to which data is regarded as true and credible
- **Believability**
 - Completeness
 - Accessibility
 - Consistency
227. 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**
 - It is in 1NF and each non key attribute is dependent upon a single column of composite primary key
 - It is in 1NF and Primary key is composite
228. ____ 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
229. Web scrapping is a process of applying ____ techniques to the web
- **Screen scrapping**
 - Data scrapping
 - Text scrapping
 - Meta scrapping
230. In which class of aggregates AVERAGE function can be placed?
- **Algebraic**
 - Distributive
 - Associative
 - Holistic
231. Which of the following is not an "Orr's Law of Data Quality"?
- "Data that is not used cannot be correct!"

AL-JÛÑÂÎD TÊCH

- “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!”
232. The ___ operator proves useful in more complex metrics applicable to the dimensions of timeliness and accessibility
- Max
 - Min
 - Min or Max
 - None
233. Which is not a/an Data Quality Validation Technique?
- Consistency Integrity
 - Referential Integrity
 - Attribute Domain
 - Using Data Quality Rules
234. Assume a company with a multi-million row customer table i.e. n rows. Checking for Referential Integrity (RI) using a naïve approach would take ___ time.
- O(n)
 - O(1)
 - O(log n)
 - None
235. ___ breaks a table into multiple tables based upon common column values
- Horizontal splitting
 - Vertical splitting
 - Both
 - None of these
236. Companies collect and record their own operational data, but at the same time they also use reference data obtained from ___ sources such as codes, prices etc.
- Operational
 - None
 - Internal
 - External
237. Ad-hoc access means to run such queries which are known already
- True
 - False
238. Relational databases allow you to navigate the data in _____ that is appropriate using the primary, foreign key structure within the data model.
- Only One direction
 - Any direction
 - Two direction
 - None

AL-JÛÑÂÎD TÊCH

239. DSS queries do not involve a primary key
- True
 - False
240. The need to synchronize data upon update is called
- Data Manipulation
 - Data Replication
 - Data Coherency
 - Data imitation
241. Taken jointly, the extract programs or naturally evolving systems formed a spider web, also known as
- Distributed Systems Architecture
 - Legacy Systems Architecture
 - Online Systems Architecture
 - Intranet Systems Architecture
242. Node of a B-Tree is stored in memory block and traversing a B-tree involves ___ page faults
- $O(n)$
 - $O(n^2)$
 - $O(n \log n)$
 - $O(\log n)$
243. Which statement is true for De-Normalization?
- Redundant data is a performance liability at query time, but is a performance benefit at update time
 - Redundant data is a performance liability at both query time and update time
 - Redundant data is a performance benefit at both query time and update time
 - Redundant data is a performance benefit at query time, but is a performance liability at update time
244. De-normalization normally speeds up
- Data Retrieval
 - Data modification
 - Development cycle
 - Data replication
245. In horizontal splitting, we split a relation into multiple tables on the basis of
- Common column values
 - Common row values
 - Different index values
 - Value resulted by ad-hoc query
246. 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:
- Owner's perspective
 - Customer's perspective
 - Decision Maker's perspective

AL-JÛÑÂÎD TÊCH

- Employee's Perspective
247. A data warehouse may include
- Legacy systems
 - Only internal data sources
 - Privacy restrictions
 - Small data mart
248. Multidimensional databases typically use proprietary ___ format to store pre-summarized cube structures
- File
 - Application
 - Aggregate
 - Database
249. All data is ___ of something real
- I An abstraction
 - II A representation
250. Which of the following option is true?
- I only
 - II only
 - Both I and II
 - None