# AL-JUNAID TECH INSTITUTE

| | | |
|---|---|---|
| Virtual University | CS614-Data Ware House<br>(Solved MCQ's)<br>LECTURE FROM<br>(23 to 45) | VU LMS HELP vulmshelp.com |
| Junaidfazal08@gmail.com<br>Bc190202640@vu.edu.pk | FOR MORE VISIT<br>VULMSHELP.COME | **JUNAID MALIK**<br>**0304-1659294** |

# AL-JUNAID TECH INSTITUTE

1. Bill Inmon argues that requirements are well understood only after
    a. They are documented
    b. They extracted and verified
    **c. Data warehouse is populated page 285**
    d. All the schemas are defined
2. Spiral model is _____
    **a. Sequence of waterfall model page 284**
    b. Risk oriented model
    c. An iterative model
    d. All of the given options
3. Waterfall is a/an_____model
    a. Iterative
    **b. Simple linear sequential page 284**
    c. Object oriented
    d. Rapid development
4. An effective user education program includes, among others, the following guideline(s);
    a. Understand the audience, don't overwhelm
    b. Train after delivery of data and analytic applications
    c. Postpone education, if DWH not ready
    **d. All of the given options page 308**
5. One of the drawbacks of waterfall model is that:
    a. Customers can not review the product during development
    b. It does not work when the resources are limited
    c. It does not define the project timeline/schedule
    **d. All of the given options page 284**
6. As per Bill Inmon, a data warehouse, in contrast with classical applicationsis:
    **a. Data driven page 285**
    b. Resource driven
    c. Requirement driven
    d. Time sensitive
7. Horizontally wide data means:
    **a. Dataset has large no. of attributes page 330**
    b. Dataset has large no. of records
    c. Dataset has attribute skews
    d. Dataset has partitioning skews
8. In context of requirement definition phase of Kimball's DWH developmentapproach, is positioned as a findings review and prioritization meeting.
    a. System analysis
    b. Scope definition
    c. Requirement configuration

**d. Requirements wrap-up presentation page 297**

9. In analytical application development phase, we follow standards for:
   a. Naming conventions
   b. Standard for calculations
   c. Standard for libraries
   **d. All of the given options page 307**

10. In lifecycle data track, we begin with translating the requirements into dimensional model, which then transforms into_____.
   **a. Physical structure page 290**
   b. Logical structure
   c. Conceptual structure
   d. System structure

11. Technical architecture design supports the communication about technical requirements:
   I. Within the team
   II. Upward to management
   III. Outward to vendors
   a. (I) Only
   b. (II) Only
   c. (III) Only
   **d. (I), (II) and (III) page 300**

12. A top down implementation approach is useful when
   **a. Technology is mature and well understood page 283**
   b. Organization can not implement latest technologies
   c. Business objectives are unclear
   d. Problem to be solved are not well documented

13. Which of the following is NOT one of the possible pitfalls in DWH LifeCycle & Development?
   a. Not having multiple servers
   b. Low priority OLAP Cube Construction
   c. Improper documentation
   **d. None of the given options page 312, 313, 314**

14. Goal driven approach of data warehouse development was result of _____ work.
   a. Bill Inmon
   b. Ralph Kimball
   **c. Bohnlein and Ulbrich-vom page 285**
   d. Westerman

15. Which of the following is/are included in the list of Top-7 key steps forsmooth DWH implementation?
   a. Consider handing-off project management

b. Assign significant resources for ETL
c. Be a diplomat NOT a technologist
**d. All of the given options page 318, 319**

16. A typical cycle of implementing the change in DWH comprises of thesequence:
a. Production -> QA -> Development
**b. Development -> QA -> Production page 314**
c. Development -> Production-> QA
d. Production-> Development -> QA

17. In_____phase of a fundamental data warehouse life cycle model, a working model of data warehouse is deployed for a selective set of users
a. Design
**b. Prototype page 287**
c. Deployment
d. Operation

18. In context of requirement definition phase in Kimball's DWH developmentapproach, activities like debriefing, documentation, and prioritization are considered to be the part of
a. Requirement preplanning
b. Business requirements collection
**c. Post collection page 294**
d. None of the given options

19. Which of the following is NOT one of the three parallel tracks in Kimballsapproach?
c. Lifecycle technology track
d. Lifecycle data track
e. Lifecycle analytical applications track
**d. Lifecycle maintenance track page 299**

19. Normally the term "DWH face to the business user" refers to:
a. Lifecycle technology track
b. Lifecycle data track
**c. Lifecycle analytical applications track page 306**
d. Lifecycle maintenance track

20. In Four Cell Quadrant Technique. The quadrant's vertical axis refers to:
a. Scope
b. Feasibility
c. Resources available
**d. The potential impact or value to the business page 297**

21. Improper documentation results the problem(s) like:
a. Maintenance issue
b. New developers unable to configure already existing code
c. Lot of time required for enhancing the code
**d. All of the given options**

22. A bottom-up implementation approach is useful when

a. Technology is mature and well understood
b. Organization can not implement latest technologies
c. The business objectives are very much clear
**d. All the requirements are well documented page 283**

23. As per Kimball,_____is the main operational process
a. Requirement extraction
b. Goal design
**c. Business process page 285**
d. Schema design

24. Which of the following is NOT one of the top-10 mistakes that should beavoided during DWH development?
a. Not interacting directly with end user
**b. Not being an accommodating person page 316, 317**
c. Isolating IT support people from business users
d. Training the users with dummy data and considering it success

25. The_____phase of fundamental data warehouse life cycle modelincludes data warehouse daily maintenance activities
a. Deployment
**b. Operation page 288**
c. Enhancement
d. Maintenance

26. In context of the most fundamental data warehouse life cycle model, whichof the following is NOT one of the data warehouse design activities?
a. End-user interviews and re-interviews
b. Source system cataloguing
c. Definition of key performance indicators
**d. System vision development page 287**

27. Which of the following is NOT one of the methodologies for Data Warehouse project development?
a. Goal Driven
b. Data Driven
c. User Driven
**d. System Driven page 283**

28. In context of analytical applications track, the application developmentactivity can begin after:
a. Database design is complete
b. A subset of historical data has been loaded
c. The data access tools and metadata are installed
**d. All of the given options page 307**

29. Waterfall model is appropriate when
a. When the budget is low
b. When the deadline is strict

c. When resources are limited

**d. Requirements are clearly defined page 284**

30. Users do not care, how advance the front end of your DWH is, what theycare is that:
    a. Tables should be properly Denormalized
    b. Proper partitioning technique should be used
    c. At least star or snowflake schema should be implemented
    **d. They should get information in timely manner and the way theywant page 314**

31. Which of the following activity executes parallel with all other activities inKimball's DWH development approach?
    a. Requirement elicitation
    b. Project planning
    **c. Project management My Point of View**
    d. Deployment

32. Which of the following is the most ignored step during data warehouse development
    a. The requirement verification
    b. The vision definition
    c. Schema validation
    **d. Success criteria development My Point of View**

33. Which of the following is NOT one of the activities of "Maintenance andGrowth phase in Kimball's DWH development approach?
    a. Education
    b. Technical Education
    c. Program Support
    **d. Interface Deployment page 309**

34. In_____phase of Kimballs approach, we identify the componentsneeded now and in future.
    a. Requirement definition
    **b. Architectural design page 300**
    c. Product development
    d. Analytical application development

35. Implementation of a data warehouse requires_____activities
    a. Highly integrated
    b. Loosely integrated
    c. Tightly decoupled
    **d. None of the given page 289**

36. Which of the following activity/activities is/are part of project planningphase in Kimballs DWH development approach?
    - ❖ Obtain resources
    - ❖ Establish the preliminary scope and justification

- ❖ Assess organization's readiness for a data warehouse initiative

**d. All of the given options page 290**

37._says that as far as company goals are concerned, the entirecompany pursues in the same direction

**a. Westerman page 285**

b. Bill Inmon

c. Kimball

d. Bohnlein

38.A dense index, if fits into memory, costs only_____disk I/O access to locate a record by given key.

- ➢ **One PG # 223**
- ➢ Two
- ➢ lg (n)
- ➢ n

39._, if fits into memory, costs only one disk I/O access to locate a record by given key.

- ➢ **A Dense Index PG # 223**
- ➢ A Sparse Index
- ➢ An Inverted Index
- ➢ None of These

40.__, if too big and doesn"t fit into the memory will be expensive when used to find a record bygiven key.

- ➢ **A Dense Index PG # 223**
- ➢ A Sparse Index
- ➢ An Inverted Index
- ➢ None of these

41.__operator is conservative in that it assigns to the dimension an aggregate value no higher than the value ofits weakest data quality indicator.

- ➢ **The Min PG # 188**
- ➢ The Simple Ratio
- ➢ The Weighted Average
- ➢ None of these

42.Parallel execution dramatically reduces response time for_____operations on large databases typicallyassociated with Decision Support Systems (DSS) and data warehouses.

- **Data-Intensive** **PG # 201**
- Quality-Intensive
- Analysis-Intensive

43. Data dependencies between different phases of computation introduce synchronization requirements that force ___execution.

- **Sequential** **PG # 204**
- Parallel
- Analogous
- Interactive Parallel

44. Maintaining locking consistency over all nodes can become a problem in large clusters. This is the disadvantageof

- **Shared Disk Approach** **PG # 209**
- Local Memory Approach
- Distributed Memory Approach

45. Performance is dictated by the_____stage in the pipeline.

- **Slowest** **PG # 217**
- Fastest
- Smallest
- Largest

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

- **Inverted Index** **PG # 232**
- OLTP
- OLAP
- DSS

47. During business hours, most_____systems should probably not use parallel execution.

- **OLTP** **PG # 206**
- OLAP
- DSS
- Data Mining

48. There are many variants of the traditional nested-loop join. If there is an index and that index is exploited, then itis called

- Naive nested-loop join
- **Index nested-loop join PG # 243**
- Temporary index nested-loop join
- None of these

49. There are many variants of the traditional nested-loop join. If the index is built as part of the query plan andsubsequently dropped, it is called
    - Naive nested-loop join
    - Index nested-loop join
    - **Temporary index nested-loop join PG # 243**
    - None of these
50. There are many variants of the traditional nested-loop join. When the entire table is scanned it is called
    - **Naive nested-loop join PG # 243**
    - Index nested-loop join
    - Temporary index nested-loop join
    - None of these
51. If someone told you that he had a good model to predict customer usage, the first thing you might try would be toask him to apply his model to your customer_____, where you already knew the answer.
    - **Base**
    - Drive
    - File
    - Log
52. Data mining is a/an_____approach, where browsing through data using data mining techniques may revealsomething that might be of interest to the user as information that was unknown previously.
    - **Exploratory PG # 249**
    - Non-Exploratory
    - Computer Science
53. Data mining uses_____algorithms to discover patterns and regularities in data.
    - **Statistical PG # 251**
    - Mathematical
    - Computational
54. Classification consists of examining the properties of a newly presented

observation and assigning it to apredefined____.

- **Class PG # 259**
- Object
- Container
- Subject

55. As opposed to the outcome of classification, estimation deal with_____valued outcome.

- **Continuous PG # 260**
- Discrete
- Isolated
- Distinct

67. is the technique in which existing heterogeneous segments are reshuffled, relocated into homogeneoussegments.

- **Clustering PG # 264**
- Aggregation
- Segmentation
- Partitioning

68. Giving the least time to_____can prove suicidal to the DWH project.

- OLAP
- De-normalization
- **ETL PG # 313**
- None of these

69. In DWH project, it is assured that_____environment is similar to the production environment

- Designing
- **Development PG # 314**
- Analysis
- Implementation

70. The application development quality-assurance activities cannot be completed until the data is_____.

- **Stabilized PG # 308**
- Identified
- Finalized
- Computerized

71. Many data warehouse project teams waste enormous amounts of time searching in vain for a ___.

- **Silver Bullet PG # 315**
- Golden Bullet
- Suitable Hardware
- Compatible Product

72. Focusing on data warehouse delivery only often end up_____.

- **Rebuilding PG # 315**
- Success
- Good Stable Product
- None of these

73. Investing years in architecture and forgetting the primary purpose of solving business problems, results ininefficient application. This is the example of_mistake.

- Extreme Technology Design
- Extreme Architecture Design
- **None of these**

74.___Division is cotton hub of Punjab.

- Lahore
- Faisalabad
- **Multan**
- Bahawalpur

75.in agriculture extension is that pest population beyond which the benefit of spraying outweighs itscost.

- None of these
- Profit Threshold Level
- **Economic Threshold Level PG # 332**
- Medicine Threshold Level

76.___is a process which involves gathering of information about column through execution of certainqueries with intention to identify erroneous records.

- **Data profiling PG # 439**
- Data Anomaly Detection
- Record Duplicate Detection
- None of these

77. Execution can be completed successfully or it may be stopped due to some error. In case of successful completionof execution all the transactions will be ___

- **Committed to the database PG # 419**
- Rolled back

78. If some error occurs, execution will be terminated abnormally and all transactions will be rolled back. In this casewhen we will access the database we will find it in the state that was

before the_____.

- **Execution of package PG # 419**
- Creation of package
- Connection of package

79. Execution can be completed successfully or it may be stopped due to some error. If some error occurs, executionwill be terminated abnormally and all transactions will be ___

- Committed to the database
- **Rolled back PG # 419**

80. To identify the degree of transformation required we need to perform_____.

- **Data Profiling PG # 437**
- Data Anomaly Detection
- Data Cleansing
- None of The Given

81. To identify the_____required we need to perform data profiling

- **Degree of Transformation PG # 437**
- Complexity
- Cost
- Time

82. To judge effectiveness we perform data profiling twice.

- One before Extraction and the other after Extraction
- **One before Transformation and the other after Transformation PG # 441**
- One before Loading and the other after Loading

83. If the dates are missing, we must need to consult_____.

- **Golden Copy PG # 456**
- Default System Date
- Silver Copy
- None of the given

84. After implementing Change Data Capture, the advantage we have is that, data is able to be integrated andtransformed _____

- **In-flight PG # 152**
- Off-flight
- Stored Data
- Over-flight

85. All data is_____of something real.

I An Abstraction
II A Representation

Which of the following option is true?

- **I Only PG # 180**
- II Only
- Both I & II
- None of I & II

86. In the Information Age, the_____learning organization is at a distinct disadvantage. This term means"impaired functioning

- Functional
- **Dysfunctional PG #181**
- Purposeful
- Serviceable

87. Many DW projects do not deliver to full potential because they treat data quality as a one-time undertaking as partof UAT. Here UAT stands for

- **User Acceptance Testing PG # 193**
- Uninterrupted Availability of Testing
- Universal Acceptance Test
- Universal Applied Test

88. NUMA stands for _____

- **Non-uniform Memory Access PG # 206**
- Non-updateable Memory Architecture
- New Universal Memory Architecture

89. Parallelism can_____system performance on over-utilized systems or systems with small I/O bandwidth.

- **Reduce PG # 202**
- Enhance
- Maintain
- Boost

90. Data mining evolve as a mechanism to cater the limitations of_____systems to deal massive data sets withhigh dimensionality, new data types, multiple heterogeneous data resources etc.

- **OLTP PG # 254**
- OLAP

- DSS
- DWH

91. In contrast to data mining, statistics is_____driven.

- **Assumption PG # 255**
- Knowledge
- Discovery
- Database

92. A__implementation approach is generally useful for projects where the technology is mature and wellunderstood, as well as where the business problems that must be solved are clear and well understood.

- **Top Down PG # 283**
- Bottom Up
- Waterfall
- Spiral

93. Implementing a data warehouse requires_____integrated activities.

- Loosely
- **Tightly PG # 289**
- Slackly
- Lethargically

94. The Kimball"s iterative data warehouse development approach drew on decades of experience to develop the____.

- OLAP Dimension
- Business Definition Lifecycle
- **Business Dimensional Lifecycle PG # 289**
- Data Warehouse Dimension

95. Pipeline parallelism focuses on increasing throughput of task execution, NOT on _____sub-task executiontime.

- Increasing
- **Decreasing PG # 215**
- Maintaining
- None of these

96. Pipeline parallelism focuses on increasing_____of task execution.

- **Throughput PG # 215**
- Non I/O Portion
- I/O Speed
- None of these

97. One needs to slot the alternative tools into categories that allow for meaningful comparison in order to____.

- **Evaluate Tools PG # 315**
- Reduce Cost
- None of these

98. Pakistan is one of the five major_____countries in the world.

- **Cotton-growing PG # 330**
- Rice-growing
- Weapon Producing

99.____is a systematic field sampling process that provide field specific information on pest pressure andcrop injury.

- **Pest Scouting PG # 333**
- Soil Survey
- Seed Survey
- Water Survey

100. Data Transformation Services (DTS) provide a set of_____that lets you extract, transform, and transformationrequired we need to perform into single or multiple destinations supported by DTS connectivity.

- **Tools PG # 373**
- Documentations
- Guidelines

101. Data Transformation Services (DTS) provide a set of tools that lets you extract, transform, and consolidate datafrom disparate sources into_supported by DTS connectivity.

- Single Destination
- Multiple Destinations
- **Single or Multiple Destinations PG # 373**

102. Data Transformation Services (DTS) provide a set of_____that lets you extract, transform, and consolidatedata from disparate sources into single or multiple destinations supported by DTS connectivity.

- **Tools PG # 373**
- Documentations
- Guidelines

103. The_____measures the ratio of desired outcomes to total outcomes.

- **Simple Ratio PG # 187**
- Min Operation
- Max Operation
- Weighted Average

104. In 1972 the Mitsubishi Shipyards in Kobe developed a technique in which customer

wants were linked to productspecifications via a matrix format. This technique is known today as:

- The Matrix of Quality
- **The House of Quality PG # 194**
- The Base Structure of Quality
- None of these

105. ___improve the overall data design and use data standards.

- Process Improvement
- System Improvement
- Policy & Procedure Improvement
- **Data Design Improvement PG # 196**

106. Which is the least appropriate join operation for Pipeline parallelism?

- Inner Join
- Inner Join
- **Sort-Merge Join**
- Hash Join

107. It must be ensured that, there are enough computing resources, Query-coordinator is very fast as compared toquery servers, Work done in each partition almost same to avoid performance bottlenecks

- To get a speed-up of N with M partitions
- To get a speed-up of N with $N^2$ partitions
- **To get a speed-up of N with N partitions PG # 213**
- To get a speed-up of N with N/2 partitions

108. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by_____tools typical of decision support systems.

- Introspective
- Intuitive
- Reminiscent
- **Retrospective**

109. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided byretrospective tools typical of___.

- **Decision Support Systems**
- OLTP
- OLAP
- Initial Data Mining Systems

110. The most recent attack is the_____attack on the cotton crop during 2003-04, resulting in a loss of nearly 0.5million bales.

- Cotton Worm
- **Boll Worm PG # 333**
- Purple Worm
- Blue Worm

111. Data recorded by pest scouts consists of two parts:

- **Static and Dynamic PG # 342**
- Valid and Invalid
- Volatile and Non-Volatile

112. DTS allows us to connect through any data source or destination that is supported by _____

- **OLE DB PG # 373**
- OLAP
- OLTP
- Data Warehouse

113. Merging information is one of the major types of _____

- **Transformation PG # 152 , 153**
- Extraction
- Loading
- None of these

114. The goal of_____is to look at as few block as possible to find the matching records.

- **Indexing PG # 222**
- Partitioning
- Joining

115. If every key in the data file is represented in the index file then index is

- **Dense Index PG # 223**
- Sparse Index
- Inverted Index
- None of these

116. ___means meeting customer‟s needs, not necessarily exceeding them.

- **Quality PG # 180**
- Marketing
- DSS
- OLAP

117. The purpose of the House of Quality technique is to reduce_____types of risk

- **Two PG # 194 , 195**
- Three
- Four
- All

118. Majority of data warehouse projects fail due to the complexity of the_____

- **Development Process PG # 283**
- Analytical Process of Cube
- Query Complexity
- Index Complexity

119. For a DWH project, the key requirement are_____and product experience.

- Tools
- **Industry PG # 320**
- Software
- None of these

120. The goal of_____is to look at as few blocks as possible to find the matching records(s).

- **Indexing PG # 222**
- Partitioning
- Joining

121. After performing most of the transformation and cleansing steps, especially after having cleaned single-sourceerror and conflicting representations, we perform___task.

- **Duplicate Elimination PG # 165**
- Duplicate Identification
- Duplicate Classification
- Duplicate Categorization

122._____improve the functional processes used to create, manage, access, and use data.

- **Process Improvement PG # 196**
- System Improvement
- Policy & Procedure Improvement
- Data Design Improvement

123. Non uniform distribution, when the data is distributed across the processors, is called _____.

- **Skew in Partition PG # 218**
- Pipeline Distribution
- Distributed Distribution
- Uncontrolled Distribution

124. In nested-loop join case, if there are „M" rows in outer table and „N" rows in inner table,

time complexity is

- O (M log N)
- O (M log N)
- **O (MN) PG # 240**
- O ($M^N$)

125. There are different DWH implementation strategies, Kimball"s Approach for data warehouse implementationis

- Data-Driven
- **Goal-Driven PG # 289**
- User-Driven
- None of these

126. If w is the window size and n is the size of data set, then the complexity of merging phase in BSN methodis__

- O (n)
- O (w)
- **O (w n) PG # 171**
- O (w log n)

127. Within the data warehousing field, data_____is applied especially when several databases are merged.

- Extraction
- Loading
- **Cleansing PG # 168**
- Join

128. Every operation cannot be parallelized, there are some preconditions and one of them is

- **The operations to be parallelized can be implemented independent of each other.**
  **PG # 201**
- The operations to be parallelized can be implemented dependent on each other.
- The operation to be parallelized has dependent sub-operations.
- None of these

129. As per Kimball,_____is the main operational process

- Requirement extraction
- Goal design
- **Business process PG # 285**
- Schema design

130. In context of data parallelism, the work done by query processor should be:

- Almost zero

- **Maximum**
- Pipelined
- Filtered across partitions

131. "More resources means proportionally less time for given amount of data". The statement refers to:

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

132. "If resources increase in proportion to increase in data size, time is constant". The statementrefers to:

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

133. Waterfall is a/an_____model.

- Iterative
- **Simple linear sequential**
- Object Oriented
- Rapid development

134. Spiral model is _____

- Sequence of waterfall model
- Risk oriented model
- An iterative model
- **All of the given options**

135. In contrast to statistics, data mining is_____driven.

- Assumption
- **Knowledge PG # 255**
- Discovery
- Database

136. In the context of Business Development Lifecycle (Kimball's approach), the first task intechnology track is_____Technical

- **Architecture Design PG # 299**
- Requirement Specification Development
- Requirement Analysis
- Lifecycle Model Selection

137. SMP Stands for_____.

- **Symmetric multi-processors PG # 2**
- **Sufficient multi-processors**

138. Identify the TRUE statement about Hypertext Transfer Protocol (HTTP).

- **HTTP is stateless protocol PG # 364**
- HTTP is not a word wide web protocol
- HTTP is used to maintain sessions
- HTTP is message routing protocol

139. ____ contribute(s) to an under-utilization of valuable and expensive historicaldata, and inevitably results in a limited capability to provide decision support and analysis.

- **The lack of data integration and standardization PG # 330**
- Less number of frequent updates
- Minimum aggregation level
- Low cube cardinality

140. For a given data set, to get a local view in un-supervised learning we use

- One-way Clustering
- **Bi-clustering PG # 271**
- Pearson correlation
- Euclidean distance

141. One-way and Two-way clustering are types of

- Supervised
- Semi-Supervised
- **Un-Supervised PG # 271**
- Reinforcement

142. If we apply Run Length Encoding on the input "111100001111", the output will be.

- **14#04#14 PG # 234**
- 41#40#41
- 18#04
- 81#40

143. PTCL is one of the examples of the following data warehouse organization

- **Telecommunications PG # 323**
- Financial service/insurance
- Transportation
- Government

144 ____ can be placed in front of our enterprise's Web servers to help them offload requests for frequently accessed content.

- **Reverse Proxy PG # 369**
- Forward Proxy

145. ____ a small piece of information generated by the Web server and stored on the client.

- **Cookie PG # 359**

146. In context of web warehousing, which of the following is NOT one of the way to identify thesession?

- Using Transient Cookies
- Using Time-contiguous Log Entries
- Using HTTP's secure sockets layer (SSL)
- **Using Simple Session Protocol (SSP) PG #364**

147. The ith bit is set to 1 if the ith row of the base table has the value for the indexed column.This statement refers to:

- Inverted index
- **Bitmap index PG233**
- Cluster index
- join index

148. Which of the following is NOT one of the issues of Clickstream data?

- Identifying the visitor origin
- Identifying the session
- Identifying the visitor
- **Identify the domain server PG # 363**

149. Which of the following is/are drawback(s) of traditional web searches?

- Limited to keyword-based matching
- Cannot distinguish between the contexts in which a link is used
- Coupling of files has to be done manually
- **All of the given options PG # 351**

150. Classification consist of examining the properties of a newly presented observation and assigning it to a predefined_____

- Object
- Class
- Container
- Subjects

151. In context of clustering the term “distance” means

- The relation of a record with corresponding recordsin child table
- None of these
- The difference between the primary keys of two records
- Similarity dissimilarity of record

152. In contrast to data mining statistics is_____driven

- Discovery
- Knowledge
- Database
- Assumption

153._____is the technique in which existing heterogeneous segmentsare reshuffled, relocated into homogenous segments.

- Clustering
- Partitioning
- Aggregation
- Segmentation

154. In context of data mining definition, the term "Value" means.

- Importance of hidden pattern discovered
- The primary key of table
- The index location of record
- Numerical or string measure assigned to an attributes

155. As oppose to the outcome of classification, estimation deal with _____valuedoutcomes

- Discrete
- Continuous
- Isolated
- distinct

156._____incorporates the con cept 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

157. The extent to which data is in appropriate languages, symbols and units, andthe definitions are clear is known as_____.

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

158. The degree to which values are present in the attributes that require them isknown as_____.

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

159. The _____dimension represents data correctness.

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

160. In B-tree index, the lowest level index blocks are called leaf blocks, andthese 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

161. Data is the_____on which a Data Warehouse (DWH) runs.

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

162. Mining Multi dimensional databases allow users to:

- Categorize the data
- Summarize the data
- Analyze the data
- All of the given

163. 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**

164. Which of the following is not an activity of Data Quality Analysis Project?

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

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

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

166. 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

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

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

168. 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**

169. Which of the following is NOT one of the parallel hardware architectures?
- Symmetric Multi-Processing
- Massively Parallel Processing
- Non-uniform Memory Access
- **Shared Memory**

170. Two interesting examples of quality dimensions that can make use of themin operator are_____.
- Believability and Consistency
- **Believability and appropriate amount of data**
- Believability and Redundancy
- Reliability and appropriate amount of data

171. As the number of processors increase the speedup should also increase. Thus we should have linear speedup. Which of the following is NOT one ofthe barriers to achieve this linear speed-up?
- Amdahl Law
- Start-up
- **No Interference**
- Skew

172. 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

173. ._____lists each term in the collection only once and then shows a list of all the documents that contain the given term.
- **Inverted index**
- Bitmap index
- Cluster index
- Join index

174. . 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)

175. _is the degree to which data accurately reflects the real-world object that the data represents
- Realistic data quality
- Strong data quality
- Intrinsic data quality
- Weak data quality

176. 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

177. 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**

178. "More resources means proportionally less time for given amount of data" that statement refers to

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

179. 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**

180. "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

181. 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

182. _is the extent to which data is regarded as true and credible.

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

183. Which is not a/an data quality validation technique?

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

184. 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**

185. _is known as state of being only one of its kind or being without anequal or parallel.
- Completeness
- **Uniqueness**
- Accessibility
- Consistency

186. Which is not a/an characteristics of data quality?
- **Reliability**
- Uniqueness
- Accessibility
- Consistency

187. 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

188. One of the main reasons for the failure of DWH deployment is _____
- **Data quality**
- Data integrity
- Data duplication
- Data anomaly

189. 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.
- **Max**
- Min
- Min and Max
- None of given

190._____is making all efforts to increase effectiveness in meeting and deficiency in meeting except customer expectations
- Quality assurance
- **Quality improvement**
- Quality maintenance
- Quality establishment

191. 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

192. 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

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

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

194. _ refers to "Parallel execution of single data operation acrossmultiple partitions of data"

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

195. _in a database or data warehouse has no actual value, it only haspotential value.

- **Data**
- Entity
- Flat tables
- Data marts

196. 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**

197. A join is identified by multiple tables in the_____clause

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

198. _____index stores first value in each block in the sequential file and a pointer to the block.

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

199. ___is a/an measure of how current or up to date the data is

- **Timeliness page 185**
- b. Completeness

- Accessibility
- Consistency

200. In context of data parallelism, the work done by query processor should be:

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

201. In context of joining tables, the join condition is specified in_____clause

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

202. 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

203. _index uses even less space than _____index, but the blockhas to be searched, even for unsuccessful searches.

- Dense, sparse
- **Sparse, dense**
- Dense, inverted
- Sparse, inverted

204. _is the degree of utility and value the data has to support theenterprise processes that enable accomplishing enterprise objectives.

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

205. _is a system of activities that assures conformance of product topre-established requirements.

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

206. 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

207. In context of bitmap index, the length of the bit vector is:

- The possible number of domain values in correspondingfield

(column)

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

208. The_____operator proves useful in more complex metric applicableto the dimensions of timeliness and accessibility.
    - **Max**
    - Min
    - Min and Max
    - None of given

209. In nested-loop join case, if there are 'M' rows in outertable and 'N' rows in inner table, time complexity is
    - O (M log N)
    - O (log MN)
    - **O (MN)**
    - O (M + N)

210. Assume a company with a multi-million row customer table i.e. nrows Checking for Referential Integrity (RI), using anaïve approach would take _____time.
    - **O(n)**
    - O(1)
    - O(log n)
    - None of the given

211. Bitmap index is appropriate for:
    - **Low cardinality data**
    - High cardinality data
    - Clustered data
    - Aggregated data

212. For a given data set, to get a global view in un-supervised learning we use
    - **One-way Clustering (Page 271)**
    - Bi-clustering
    - Pearson correlation
    - Euclidean distance

213. _____ is the technique in which existing heterogeneous segments are reshuffled, relocated into homogeneous segments.
- **Clustering**
- Aggregation
- Segmentation
- Partitioning

214. In data mining, initially you _____ what you are lo(Correct)ing for.
- Know
- **Don't know**
- May or may not know
- None of the given options

215. Bill Inmon argues that requirements are well understood only after
- They are documented

- **They extracted and verified**
- Data warehouse is populated
- All the schemas are defined

216. In _____learning you don't know the number of clusters and no idea about their attributes.
- Supervised learning
- **Unsupervised learning**
- Multi Dimension modeling
- None of the given options

217. Identify the TRUE statement:
- **The data value increases as volume decreases**
- The data value decreases as the volume decreases
- The data value is independent of data volume
- All of the given options

218. As per Kimball, _____ is the main operational process
- Requirement extraction
- Goal design
- **Business process**
- Schema design

219. In context of data mining definition, the term "value" means:
- The primary key of table
- **The index location of the record**
- Importance of hidden patterns discovered
- Numerical or string measure assigned to an attribute

220. Data mining is all about:
- **Knowledge discovery in database**
- Finding hidden patterns in data
- Finding relationships in data
- All of the given options

221. Normally the input data structure (a database table) for a data mining algorithm:
- Has limited amount of data
- Has limited amount of relations
- **Has more number of attributes than records**
- Has more number of records than attributes

222. Vertically wide data means:
- Dataset has large no. of attributes
- **Dataset has large no. of records**
- Dataset has attribute skews
- Dataset has partitioning skews

223. In _____ phase of kimballs approach, we identify the components needed now and in future.
- Requirement definition
- **Architectural design**
- Product development
- Analytical application development

224. Implementation of a data warehouse requires _____ activities
- Highly integrated
- Loosely integrated
- **Tightly decoupled**

- None of the given

225. In Four Cell Quadrant Technique, The quadrant's vertical axis refers to:
- Scope
- Feasibility
- Resources available
- **The potential impact or value to the business**

226. The phase-3 i.e., "Rollout and Maintenance" phase in "12-steps data warehouse implementation approach" of Shaku Atre comprises of _____ activity.
- On Line Analytical Processing
- Construct Metadata Repository
- **Deployment & System Management**
- **None of the given options**

227. Which of the following is NOT one of the examples of dynamic attributes?
- Daily Sale
- **Date of Birth**
- Air pressure
- None of the given options

228. Users do not care, how advance the front end of your DWH is, what they care is that:
- Tables should be properly denormalized
- Proper partitioning technique should be used
- At least star or snow flake schema should be implemented
- **They should get information in timely manner and the way they want**

229. Which of the following is NOT one of the top-10 mistakes that should be avoided during DWH development?
- Not interacting directly with end user
- **Not being an accommodating person**
- Isolating IT support people from business users
- Training the users with dummy data and considering it success

230. Normally it is recommended to have
- **Interference while having different database environments on a single server**

231. Identify the TRUE statement:
- **Clustering is unsupervised learning and classification is supervised learning**
- Both clustering and Classification are supervised Learning
- Clustering is supervised learning and classification is unsupervised learning
- Both clustering and Classification are unsupervised learning

232. "What means What". The phrase refers to
- **Meta Data**
- Transformed Data
- Internal Representation
- External Data

233. In context of requirement definition phase of Kimball's DWH development approach, is positioned as findings review and prioritization meeting.
- **Requirements wrap-up presentation**

234. The first step of the "12-steps data warehouse implementation approach' of Shaku Atre ts:
- **Finding user needs**
- Planning System Resources
- Finding System Scope
- Data Acquisition and Cleansing

235. Which of the following is/are example(s) of static attributes?

- **Employee Name**
- Employee Date of Birth
- Employee Blood Group
- All of given

236. In _____learning you don't know the number of clusters and no idea about their attributes.

- **Unsupervised learning**