CS614 Quiz No. 1 Solution (Jan 2021)

If found any wrong, correct it accordingly.

1. _____ incorporates the concept of product quality, process control, quality assurance, and quality improvement.
   a. **Total Quality Management page 182**
   b. Intrinsic Data Quality Management
   c. Realistic Data Quality Management
   d. Strong Data Quality Management
2. 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
3. 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
4. The _____ dimension represents data correctness.
   a. **Free-of-error page 187**
   b. Completeness
   c. Consistency
   d. Correctness
5. In B-tree index, the lowest level index blocks are called leaf blocks, and these blocks contain:
   a. NULL value to make the leaf terminal node
   b. **Every indexed data value and a corresponding ROWID page 225**
   c. Every indexed data value and pointer to next level block
   d. Every indexed data value and pointer to root block
6. Data is the _____ on which a Data Warehouse (DWH) runs.
   a. **Fuel page 179**
   b. Element
   c. Component
   d. Entity

MCS of Virtuallians Group Team

7. In context of data parallelism to get a speed-up of N with N partitions, it must be ensured that:
    a. There are enough computing resources
    b. Query-coordinator is very fast as compared to query servers
    c. Work done in each partition almost same
    d. All of the given options page 213
8. Which of the following is not an activity of Data Quality Analysis Project?
    a. "Define"
    b. "Measure"
    c. "Analyze"
    d. "Compression" page 194
9. Which of the following is not a Data Quality Validation Technique?
    a. Referential Integrity
    b. Using Data Quality Rules
    c. Data Histograming
    d. Indexes page 189
10. One of the preconditions to decide about operations to be parallelized is that
    a. Operation can be implemented independent of each other page 201
    b. Output of one operation becomes input of other
    c. Operations share same memory location
    d. Operations share same namespace
11. _____do not (typically) keep the index values in sorted order
    a. Dense index
    b. Sparse index
    c. B-Tree Index
    d. Hash Based index page 227
12. Parallelism can be exploited, if there is :
    a. Symmetric multi processors (SMP)
    b. Sufficient I/O bandwidth
    c. Underutilized or intermittently used CPUs
    d. All of the given options page 202
13. Which of the following is NOT one of the parallel hardware architectures?
    a. Symmetric Multi-Processing
    b. Massively Parallel Processing
    c. Non-uniform Memory Access
    d. Shared Memory page 206

MCS of Virtuallians Group Team

14. Two interesting examples of quality dimensions that can make use of the min operator are _____.
    a. **Believability and appropriate amount of data page 188**
    b. Believability and Consistency
    c. Believability and Redundancy
    d. Reliability and appropriate amount of data
15. 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?
    a. Amdahl Law
    b. Startup
    c. **No Interference page 219**
    d. Skew
16. In _____index, the ith bit is set to "1" if the ith row of the base table has the value for the indexed column
    a. Inverted index
    b. **Bitmap index page 233**
    c. Cluster index
    d. Join index
17. _____ lists each term in the collection only once and then shows a list of all the documents that contain the given term.
    a. **Inverted index page 232**
    b. Bitmap index
    c. Cluster index
    d. Join index
18. The exact formula for Speed-up is:
    a. **(Time on Serial Processor) / (Time on parallel processors) page 204**
    b. (Time on Serial Processor) * (Time on parallel processors)
    c. (Time on Serial Processor) + (Time on parallel processors)
    d. (Time on Serial Processor) - (Time on parallel processors)
19. _____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
    d. Weak data quality

20. 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.
    a. O(log n) page 190
    b. O(n)
    c. O(1)
    d. None of the given
21. Which of the following is NOT one of the variants of Nested-loop join?
    a. Naive nested-loop join
    b. Indexed nested-loop join
    c. Temporary index nested-loop join
    d. Binary index nested-loop join page 243
22. "More resources means proportionally less time for given amount of data" that statement refers to
    a. Scale-Up
    b. Speed-Up page 203
    c. Size-Up
    d. Over-Utilized system
23. The optimizer uses a hash join to join two tables if they are joined using an equijoin and
    a. outer table has less number of rows
    b. inner table has less number of rows
    c. cardinality of table is equal
    d. large amount of data needs to be joined page 245
24. "If resources increase in proposition to increase in data size, time is constant". The statement refers to
    a. Scale-up page 203
    b. Speed-up
    c. Size-up
    d. Over-utilized system
25. 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 _____.
    a. Defective page 180
    b. Unclear
    c. Unrefined
    d. Undefined

MCS of Virtuallians Group Team

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

a. Believability page 188

b. Completeness

c. Accessibility

d. consistency

27. 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

28. Which of the following is not an "Orr's law of data quality"?

a. Data that Is not used is cannot be correct

b. Data quality is 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

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

a. Completeness

b. Uniqueness page 185

c. Accessibility

d. Consistency

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

a. Reliability page 186

b. Uniqueness

c. Accessibility

d. Consistency

31. If every key in the data file is represented in the index file then it is called

a. Dense Index page 223

b. Sparse Index

c. Inverted Index

d. A Multi level Sparse Index

32. One of the main reasons for the failure of DWH deployment is ____

a. Data quality page 179

b. Data integrity

c. Data duplication

d. Data anomaly

33. 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.
    a. Min page 188
    b. Max
    c. Min and Max
    d. None of given
34. ____ is making all efforts to increase effectiveness in meeting and deficiency in meeting except customer expectations
    a. Quality assurance
    b. Quality improvement page 183
    c. Quality maintenance
    d. Quality establishment
35. Most DWH implantations today do not use ____ enforced by the database, but as TQM method improved overall data quality and database optimizers.
    a. Consistency integrity
    b. Referential integrity page 190
    c. Attribute domain
    d. Using data quality rules
36. 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.
    a. N*T page 215
    b. N/T
    c. N+T
    d. N-T
37. An optimized structure which is built primarily for retrieval, with update being only a secondary consideration is
    a. OLTP
    b. OLAP
    c. DSS
    d. Inverted Index page 232
38. _____ refers to "Parallel execution of single data operation across multiple partitions of data"
    a. Hardware parallelism
    b. Software parallelism
    c. Data parallelism page 212
    d. Operational parallelism

MCS of Virtuallians Group Team

39. _____ in a database or data warehouse has no actual value, it only has potential value.
    a. Data page 181
    b. Entity
    c. Flat tables
    d. Data marts

MCS of Virtuallians Group Team

40. Which of the following tasks can NOT be parallelized?
    a. Large table scans and joins
    b. Creation of large indexes
    c. Partitioned index scans
    d. None of the given options page 201

41. A join is identified by multiple tables in the _____ clause
    a. FROM page 239
    b. SELECT
    c. GROUP BY
    d. SORT BY

MCS of Virtuallians Group Team

42. _____ index stores first value in each block in the sequential file and a pointer to the block.
    a. Dense
    b. Sparse page 224
    c. B-Tree
    d. Hash

MCS of Virtuallians Group Team

43. _____ is a/an measure of how current or up to date the data is
    a. Timeliness page 185
    b. Completeness
    c. Accessibility
    d. Consistency

MCS of Virtuallians Group Team

44. In context of data parallelism, the work done by query processor should be:
    a. Almost zero
    b. Maximum
    c. Pipelined
    d. Filtered across partitions

MCS of Virtuallians Group Team

45. In context of joining tables, the join condition is specified in _____clause
    a. FORM
    b. SELECT
    c. WHERE
    d. GROUP BY

46. A _____ index, if fits in the memory, costs only one disk I/O access to locate a record given a key.
    a. **Dense page 223**
    b. Sparse
    c. B-Tree
    d. Hash

MCS of Virtuallians Group Team

47. _____ index uses even less space than _____ index, but the block has to be searched, even for unsuccessful searches.
    a. Dense, sparse
    b. **Sparse, dense page 224**
    c. Dense, inverted
    d. Sparse, inverted

MCS of Virtuallians Group Team

48. _____ is the degree of utility and value the data has to support the enterprise processes that enable accomplishing enterprise objectives.
    a. Intrinsic Data Quality
    b. **Realistic Data Quality page 181**
    c. Strong Data Quality
    d. Weak Data Quality

MCS of Virtuallians Group Team

49. _____ is a system of activities that assures conformance of product to pre-established requirements.
    a. **Quality assurance page 183**
    b. Quality improvement
    c. Quality maintenance
    d. Quality establishment

MCS of Virtuallians Group Team

50. 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
    a. Dependent
    b. **Independent page 241**
    c. Superset
    d. Subset

MCS of Virtuallians Group Team

51. In context of bitmap index, the length of the bit vector is:
    a. The possible number of domain values in corresponding field (column)
    b. **The number of records in the base table page 233**
    c. The possible number of bitmap tables formed for corresponding field (column)
    d. None of the given options

52. The _____ operator proves useful in more complex metrics applicable to the dimensions of timeliness and accessibility.
    a. Max page 188
    b. Min
    c. Min and Max
    d. None of given

MCS of Virtuallians Group Team

53. In nested-loop join case, if there are 'M' rows in outer table and 'N' rows in inner table, time complexity is
    a. O (M log N)
    b. O (log MN)
    c. O (MN) page 241
    d. O (M + N)

MCS of Virtuallians Group Team

54. 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.
    a. O(n) page 190
    b. O(1)
    c. O(log n)
    d. None of the given

MCS of Virtuallians Group Team

55. Bitmap index is appropriate for:
    a. Low cardinality data
    b. High cardinality data
    c. Clustered data
    d. Aggregated data
56.

MCS of Virtuallians Group Team

MCS of Virtuallians Group Team