بسم الله الرحمن الرحيم

In the Name of Allāh, the Most Gracious, the Most Merciful

Final-Term Papers Solved MCQS with Reference

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

- Linear
- Quadratic
- Logarithmic
- **Exponential** **PG # 15**

2.The telecommunications data warehouse is dominated by the sheer volume of data generated at the call level _____ area.

- **Subject** **PG # 35**
- Object
- Aggregate
- Detail

3.De-Normalization normally speeds up

- **Data Retrieval** **PG # 51**
- Data Modification
- Development Cycle
- Data Replication

4.ER Model can be simplified in -------- ways

- One
- **Two** **PG # 103**
- Three
- Four

There are actually two ways of "simplifying" the ER model i.e. (i) De-normalization and (ii) Dimensional Modeling.

5.Non recording facts have a disadvantage that it has

- **Lack of Information** **PG # 120**
- Redundant Information
- Repeated Information
- Normalized Information

6.Fact-less fact table is a fact table without numeric fact columns. It is used to capture relationship between _____

- **Dimensions** **PG # 121**
- Attributes
- Tables
- Facts

7.A cube is not a data warehouse, it is a _____

- **Data Mart** **PG # 131**
- Data Extraction
- Data Loading
- Query Manager

8.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

9._____, 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

10._____, if too big and doesn't fit into the memory will be expensive when used to find a record by given key.

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

11._____operator is conservative in that it assigns to the dimension an aggregate value no higher than the value of its weakest data quality indicator.

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

12.Parallel execution dramatically reduces response time for _____ operations on large databases typically associated with Decision Support Systems (DSS) and data warehouses.

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

13.Data dependencies between different phases of computation introduce synchronization requirements that force _____ execution.

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

14.Maintaining locking consistency over all nodes can become a problem in large clusters. This is the disadvantage of

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

15.Performance is dictated by the _____ stage in the pipeline.

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

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

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

17.During business hours, most _____ systems should probably not use parallel execution.

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

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

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

19.There are many variants of the traditional nested-loop join. If the index is built as part of the query plan and subsequently dropped, it is called

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

20.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

21.If someone told you that he had a good model to predict customer usage, the first thing you might try would be to ask him to apply his model to your customer _____, where you already knew the answer.

- **Base** **Click here for Reference Detail**
- Drive
- File
- Log

22.Data mining is a/an _____ approach, where browsing through data using data mining techniques may reveal something that might be of interest to the user as information that was unknown previously.

- **Exploratory** **PG # 249**
- Non-Exploratory
- Computer Science

23.Data mining uses _____ algorithms to discover patterns and regularities in data.

- **Statistical** **PG # 251**
- Mathematical
- Computational

24.Classification consists of examining the properties of a newly presented observation and assigning it to a predefined _____.

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

25.As opposed to the outcome of classification, estimation deal with _____ valued outcome.

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

26._____ is the technique in which existing heterogeneous segments are reshuffled, relocated into homogeneous segments.

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

27.Giving the least time to _____ can prove suicidal to the DWH project.

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

28.In DWH project, it is assured that _____ environment is similar to the production environment

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

29.The application development quality-assurance activities cannot be completed until the data is _____.

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

30.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

Many data warehouse project teams waste enormous amounts of time searching in vain for a silver bullet i.e. a panacea or Amratdhara

31.Focusing on data warehouse delivery only often end up _____.

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

Focusing on data warehouse delivery, architecture feels like a distraction and impediment to progress and often end up rebuilding.

32.Investing years in architecture and forgetting the primary purpose of solving business problems, results in inefficient application. This is the example of _____ mistake.

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

33._____ Division is cotton hub of Punjab.

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

Multan, a division in southern Punjab, is known as the agricultural hub of the country. Multan accounts for 70% cotton production of the country.

34._____ in agriculture extension is that pest population beyond which the benefit of spraying outweighs its cost.

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

ETL A: Economic Threshold Level in agriculture extension is that pest population beyond which the benefit of spraying outweighs its cost. It is highly infeasible and expensive to eradicate

35._____ is a process which involves gathering of information about column through execution of certain queries with intention to identify erroneous records.

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

36.If we remove the modification anomalies the table comes in _____ form

- 1NF
- 2NF
- **3NF** **PG # 47**
- 4NF

37.Modification anomalies occur in _____ form

- 1NF
- **2NF** **PG # 45**
- 3NF
- BCNF

38.Execution can be completed successfully or it may be stopped due to some error. In case of successful completion of execution all the transactions will be _____

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

39.If some error occurs, execution will be terminated abnormally and all transactions will be rolled back. In this case when 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

40.Execution can be completed successfully or it may be stopped due to some error. If some error occurs, execution will be terminated abnormally and all transactions will be _____

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

41.To identify the degree of transformation required we need to perform _____.

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

42.To identify the _____ required we need to perform data profiling

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

43.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

44.If the dates are missing we must need to consult _____.

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

45.In MOLAP physically build cubes for direct access, support is not available for _____ SQL.

- **ANSI** **PG # 78**
- Microsoft
- Oracle
- SAP

MOLAP physically builds "cubes" for direct access - usually in the proprietary file format of a multi-dimensional database (MDD) or a user defined data structure. Therefore ANSI SQL is not supported.

46._____ is the lowest level of detail or the atomic level of data stored in the warehouse.

- Aggregate
- Cube
- **Grain** **PG # 111**
- Virtual Cube

47. After implementing Change Data Capture, the advantage we have is that, data is able to be integrated and transformed _____

- **In-flight** **Click Here For Reference Detail** **PG # 152**
- Off-flight
- Stored Data
- Over-flight

Finally data is able to be integrated and transformed "in-flight". Once the update/transaction data has been pulled from the log tape, the DWH is free to re-sequence, reformat, convert, merge, summarize, etc.

48. 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

49. In the Information Age, the _____ learning organization is at a distinct disadvantage. This term means "impaired functioning

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

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

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

51.NUMA stands for _____

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

52.Parallelism can _____system performance on over-utilized systems or systems with small I/O bandwidth.

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

53.Data mining evolve as a mechanism to cater the limitations of _____ systems to deal massive data sets with high dimensionality, new data types, multiple heterogeneous data resources etc.

- **OLTP** **PG # 254**
- OLAP
- DSS
- DWH

54.In contrast to data mining, statistics is _____ driven.

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

55.A _____ implementation approach is generally useful for projects where the technology is mature and well understood, as well as where the business problems that must be solved are clear and well understood.

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

56.Implementing a data warehouse requires _____ integrated activities.

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

57.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

58.Pipeline parallelism focuses on increasing throughput of task execution, NOT on _____ sub-task execution time.

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

59.Pipeline parallelism focuses on increasing _____ of task execution.

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

60.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

61.Pakistan is one of the five major _____ countries in the world.

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

62._____is a systematic field sampling process that provide field specific information on pest pressure and crop injury.

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

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

- 1950s.
- **1960s.** **PG # 12**
- 1970s.
- 1980s.

64.Redundancy causes _____ anomalies

- **Update** **PG # 43**
- Select
- Both Update & Select
- None of these

65.Data Transformation Services (DTS) provide a set of _____ that lets you extract, transform, and transformation required we need to perform into single or multiple destinations supported by DTS connectivity.

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

66.Data Transformation Services (DTS) provide a set of tools that lets you extract, transform, and consolidate data from disparate sources into _____ supported by DTS connectivity.

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

67.Data Transformation Services (DTS) provide a set of _____ that lets you extract, transform, and consolidate data from disparate sources into single or multiple destinations supported by DTS connectivity.

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

68.In ROLAP access to information is provided via relational database using _____ standard SQL.

- **ANSI** **PG # 78**
- Microsoft
- Oracle
- SAP

69. Which of the following statement is true? 1 GB is

- **2^{30} or 10^9 bytes** **PG # 15**
- 2^{30} or 10^6 bytes
- 2^{32} or 10^9 bytes
- 2^{32} or 10^8 bytes

How Much Data is that?		
1 MB	2^{20} or 10^6 bytes	Small novel – 31/2 Disk
1 GB	2^{30} or 10^9 bytes	Paper rims that could fill the back of a pickup van
1 TB	2^{40} or 10^{12} bytes	50,000 trees chopped and converted into paper and printed
2 PB	1 PB = 2^{50} or 10^{15} bytes	Academic research libraries across the U.S.
5 EB	1 EB = 2^{60} or 10^{18} bytes	All words ever spoken by human beings

Table-2.1: Quantifying size of data

70. Which of the following statement is true? 1 PB is

- 2^{52} or 10^{13} bytes
- **2^{50} or 10^{15} bytes** **PG # 15**
- 2^{50} or 10^{10} bytes
- 2^{48} or 10^{12} bytes

71.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 lg n)
- **O (log n)** **PG # 22**

72.Normally Selectivity of query in OLTP system is

- **High** **PG # 30**
- Low
- Not measured

OLTP	DWH
Primary key used	Primary key NOT used
No concept of Primary Index	Primary index used
May use a single table	Uses multiple tables
Few rows returned	Many rows returned
High selectivity of query	Low selectivity of query
Indexing on primary key (unique)	Indexing on primary index (non-unique)

Table-4.1: Comparison of OLTP and DWH for given queries

73.Normally Selectivity of query in data warehouse is

- High
- **Low** **PG # 30**
- Not measured

74.One major goal of horizontal splitting is

- Splitting rows for exploiting parallelism
- Splitting columns for exploiting parallelism
- Splitting schema for exploiting parallelism
- **Spreading rows for exploiting parallelism.** **PG # 46**

Splitting Tables: Horizontal splitting

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

GOAL

- Spreading rows for exploiting parallelism.
- Grouping data to avoid unnecessary query load in WHERE clause.

75.Fact-less fact table is a fact table without numeric fact columns. It is used to capture relationship between _____

- **Dimensions** **PG # 121**
- Attributes
- Tables
- Facts

A Fact-less Fact Table

- "Fact -less" fact table
  - A fact table without numeric fact columns
  - Captures relationships between dimensions
  - Use a dummy fact column that always has value 1

76.The _____ measures the ratio of desired outcomes to total outcomes.

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

77.In 1972 the Mitsubishi Shipyards in Kobe developed a technique in which customer wants were linked to product specifications 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

78._____ improve the overall data design and use data standards.

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

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

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

80.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 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

81.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** **Click Here For Reference Detail**

82.The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of _____.

- **Decision Support Systems** **Click Here For Reference Detail**
- OLTP
- OLAP
- Initial Data Mining Systems

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

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

84._____ gives total view of an organization

- OLTP
- **Data warehouse** **PG # 16**
- OLAP
- Data base

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

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

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

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

87.Experience showed that for a single pass magnetic tape that scanned 100% of the records, only _____ of the records, sometimes even were actually required.

- **5%** **PG # 12**
- 30%
- 50%
- 80%

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

- **Doubles** **PG # 15**
- Triples
- Quartiles
- Remains same as previous year

89.Normalized design is likely to perform much faster than de-normalized design for queries that probe

- **Master table only** **PG # 64**
- Details tables only
- Both master and detail tables

90. Partition elimination is not possible with

- **Round-Robin** **PG # 66**
- De-normalization
- Normalization

91. ER is a logical design technique that seeks to remove the _____ in data.

- **Redundancy** **PG # 98**
- Normalization
- Anomalies

92.ER is a _____ design technique that seeks to remove the redundancy in data.

- **Logical** **PG # 98**
- Physical
- Data Dependent
- Transaction Dependent

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

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

Data merging is part of data transformation where multiple values are summarized into single summarized value.

94. The goal of _____is to look at as few block as possible to find the matching records.

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

95. 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

96. _____ means meeting customer's needs, not necessarily exceeding them.

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

97. The purpose of the House of Quality technique is to reduce _____ types of risk

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

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

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

99. For a DWH project, the key requirement are _____ and product experience.

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

100. 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** **PG # 19**
- Two Direction
- None of these

101. In _____ system, the contents change with time.

- **OLTP** **PG # 20**
- DSS
- ATM
- OLAP

102. Primary key is repeated in _____ splitting.

- Horizontal
- **Vertical** **PG # 56**

103.Geography is a good example of

- One-dimensional Hierarchy
- **Multidimensional Hierarchy** **PG # 52**
- Non-Dimensional
- Linear Hierarchy

104.Cube is a logical entity containing values of a certain fact at a certain aggregation level at _____ of a combination of dimensions.

- **An Intersection** **PG # 88**
- A Union
- A Subtraction
- A Subset

105.Pre-computed _____ can solve performance problems

- **Aggregates** **PG # 111**
- Facts
- Dimensions

106.A company has implemented data warehouse for analytical purpose. Quantity sold is stored as a fact. This quantity sold is

- **Additive Fact** **PG # 119**
- Non-Additive Fact

107.In full extraction, data is extracted completely from the source system. Therefore there is no need to keep track of changes to the _____

- **Data Source** **PG # 133**
- DWH
- Data Mart

108. The goal of _____ is to look at as few blocks as possible to find the matching records(s).

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

109. After performing most of the transformation and cleansing steps, especially after having cleaned single-source error and conflicting representations, we perform _____ task.

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

110. _____ improve the functional processes used to create, manage, access, and use data.

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

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

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

112.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)

If the outer loop executes R times and for each such execution the inner loop executes S times, then the total cost or time complexity of the nested loop is *O(RS)*.

113.There are different DWH implementation strategies, Kimball’s Approach for data warehouse implementation is

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

114.If w is the window size and n is the size of data set, then the complexity of merging phase in BSN method is_____

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

115._____ is one class of decision support environment.

- **OLAP** **PG # 30**
- OLTP
- Data Cleansing
- ETL

116.Within the data warehousing field, data _____ is applied especially when several databases are merged.

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

117.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

118.The users of data warehouse are _____

- Decision makers
- Knowledge workers
- **Both Knowledge workers and Decision makers** **PG # 18**

The users of data warehouse are knowledge workers in other words they are decision makers in the organization.

119.As per Kimball, _____ is the main operational process

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

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

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

121.“More resources means proportionally less time for given amount of data”. The statement refers to:

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

122.“If resources increase in proportion to increase in data size, time is constant”. The statement refers to:

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

123.Waterfall is a/an _____ model.

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

124.Spiral model is _____

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

125.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

126.Effects of de-normalization on database performance are

- **Unpredictable** **PG # 62**
- Predictable
- Conventional
- Unsurprising

127.OLAP is used for analytical process. For analytical processing we need

- **Multi-level aggregates** **PG # 74**
- Record level access
- Data level access
- Row level access

128.In contrast to statistics, data mining is _____ driven.

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

129.In the context of Business Development Lifecycle (Kimball's approach), the first task in technology track is _____ Technical

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

130.Multidimensional databases typically use proprietary _____ format to store pre-summarized cube structures.

- **File** **PG # 79**
- Application
- Aggregate
- Database

131.As consumers, human beings judge the quality of things during their life-time.

I Consciously

II Subconsciously

III Unconsciously

Which of the following statement is true?

- I Only
- II Only
- III Only
- **I & II Only** **PG # 179**

132.Product selection phase fall in the _____ Kimball's approach of business dimensional life cycle.

133.SMP Stands for _____.

- **Symmetric multi-processors** **PG # 202**
- Sufficient multi-processors

134.Records referring to the same entity are represented in different formats in the different data sets or are represented erroneously. Thus, duplicate records will appear in the merged database. The issue is to identify and eliminate these duplicates. The problem is known as the _____.

- **Merge/Purge Problem** **PG # 168**
- Cleansing Problem
- Transformation Problem
- Data Quality Problem

135.The users of data warehouse are knowledge workers in other words they are _____ in the organization.

- **Decision maker** **PG # 18**
- Manager
- Database Administrator
- DWH Analyst

136.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

137._____ contribute(s) to an under-utilization of valuable and expensive historical data, 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

138.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

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

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

140.Collapsing tables can be done on the _____ relationship(s)

- Only One-to-One
- Only Many-to-Many
- Only One-to-Many
- **Both One-to-One and Many-to-Many** **PG # 52**

141.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

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

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

DWH Target Organizations
• **Financial service/insurance.**
– Union Bank
– State Bank of Pakistan
• **Telecommunications.**
– UFone
– PTCL
– PAKNET
• **Transportation.**
– PIA
• **Government.**
– NADRA

143._____ 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

144._____a small piece of information generated by the Web server and stored on the client.

- **Cookie** **PG # 359**

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

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

146.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** **PG # 233**
- Cluster index
- join index

147.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**

Clickstream data has many issues.
1. Identifying the Visitor Origin
2. Identifying the Session
3. Identifying the Visitor
4. Proxy Servers
5. Browser Caches

148.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**