

CS614-Data Warehouse FINAL TERM MCQS

Prepared by: JUNAID MALIK

AL-JUNAID TECH INSTITUTE



www.vulmshelp.com



Language Courses Training Available

I'm providing paid courses in different languages within 3 Months, Certificate will be awarded after completion.

- HTML
- CSS
- JAVASCRIPT
- BOOTSTRAPS
- JQUERY
- PHP MYSQL
- NODES.JS
- REACT JS

LMS Handling Services

LMS Activities Paid Task

Assignments 95% Results

Quizes 95% Results

GDB 95% Results

For CS619 Project Feel Free To Contact With Me

Ph# 0304-1659294

Email: junaidfzal08@gmail.com

ALL answers are verified if found any mistake then Correct ACCORDINGLY

AL-JUNAID INSTITUTE GROUP

1. Classification consist of examining the properties of a newly presented observation and assigning it to a predefined _____
 - Object
 - **Class**
 - Container
 - Subjects
2. In context of clustering the term “distance” means
 - The relation of a record with corresponding records in child table
 - None of these
 - The difference between the primary keys of two records
 - **Similarity dissimilarity of record**
3. In contrast to data mining statistics is _____ driven
 - Discovery
 - Knowledge
 - Database
 - **Assumption**
4. _____ is the technique in which existing heterogeneous segments are reshuffled, relocated into homogenous segments.
 - **Clustering**
 - Partitioning
 - Aggregation
 - Segmentation
5. 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
6. As oppose to the outcome of classification, estimation deal with _____ valued outcomes
 - Discrete
 - **Continuous**
 - Isolated
 - distinct
7. _____ 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

AL-JUNAID INSTITUTE GROUP

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

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

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

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

10. The _____ dimension represents data correctness.

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

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

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

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

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

13. Mining Multi dimensional databases allow users to:

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

AL-JUNAID INSTITUTE GROUP

14. 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**
15. Which of the following is not an activity of Data Quality Analysis Project?
- "Define"
 - "Measure"
 - "Analyze"
 - **"Compression"**
16. Which of the following is not a Data Quality Validation Technique?
- a. Referential Integrity
 - b. Using Data Quality Rules
 - c. Data Histograming
 - d. Indexes**
17. One of the preconditions to decide about operations to be parallelized is that
- a. Operation can be implemented independent of each other**
 - b. Output of one operation becomes input of other
 - c. Operations share same memory location
 - d. Operations share same namespace
- 18.. _____ do not (typically) keep the index values in sorted order
- a. Dense index
 - b. Sparse index
 - c. B-Tree Index
 - d. Hash Based index**
19. 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**
- 20.. Which of the following is NOT one of the parallel hardware architectures?
- Symmetric Multi-Processing
 - Massively Parallel Processing
 - Non-uniform Memory Access
 - **Shared Memory**

AL-JUNAID INSTITUTE GROUP

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

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

22. . As the number of processors increase the speedup should also increase.

Thus we should have linear speedup. Which of the following is NOT one of the barriers to achieve this linear speed-up?

- Amdahl Law
- Start-up
- **No Interference**
- Skew

23. 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

24. . _____ 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

25. . 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)

24. _____ 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

AL-JUNAID INSTITUTE GROUP

25. 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
26. 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**
27. “More resources means proportionally less time for given amount of data” that statement refers to
- Scale-Up
 - **Speed-Up**
 - Size-Up
 - Over-Utilized system
28. 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**
29. “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
30. 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**
 - Unclear
 - Unrefined
 - Undefined

AL-JUNAID INSTITUTE GROUP

31. _____ is the extent to which data is regarded as true and credible.

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

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

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

33. 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**

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

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

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

- **Reliability**
- Uniqueness
- Accessibility
- Consistency

36. 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

37. One of the main reasons for the failure of DWH deployment is _____

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

AL-JUNAID INSTITUTE GROUP

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

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

39. _____ is making all efforts to increase effectiveness in meeting and deficiency in meeting except customer expectations

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

40. 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

41. 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

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

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

43. _____ refers to “Parallel execution of single data operation across multiple partitions of data”

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

AL-JUNAID INSTITUTE GROUP

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

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

45. 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**

46. A join is identified by multiple tables in the _____ clause

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

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

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

48. _____ is a/an measure of how current or up to date the data is

- **a. Timeliness** page 185
- b. Completeness
- c. Accessibility
- d. Consistency

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

- **a. Almost zero** Goggle
- b. Maximum
- c. Pipelined
- d. Filtered across partitions

50. In context of joining tables, the join condition is specified in _____ clause

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

AL-JUNAID INSTITUTE GROUP

51. 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

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

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

53. _____ 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

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

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

55. 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

56. 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**
- c. The possible number of bitmap tables formed for corresponding field (column)
- d. None of the given options

AL-JUNAID INSTITUTE GROUP

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

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

58. In nested-loop join case, if there are 'M' rows in outer table and 'N' rows in inner table, time complexity is

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

59 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 of the given

60. Bitmap index is appropriate for:

- **Low cardinality data**
- High cardinality data
- Clustered data
- Aggregated data

60. A data warehouse may include

- ▶ **Legacy systems**
- ▶ Only internal data sources
- ▶ Privacy restrictions
- ▶ Small data mart

61.

De-Normalization normally speeds up

- ▶ **Data Retrieval** (page no 51)

AL-JUNAID INSTITUTE GROUP

- ▶ Data Modification
- ▶ Development Cycle
- ▶ Data Replication

63.

In horizontal splitting, we split a relation into multiple tables on the basis of

- ▶ **Common Column Values** page no 54
- ▶ Common Row Values
- ▶ Different Index Values
- ▶ Value resulted by ad-hoc query

64.

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

- ▶ **File page no 79**
- ▶ Application
- ▶ Aggregate
- ▶ Database

65.

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

- ▶ **One** [page no 233]
- ▶ Two
- ▶ $\lg(n)$
- ▶ n

66.

All data is _____ of something real.

IA An Abstraction

II A Representation

Which of the following option is true?

- ▶ I Only
- ▶ II Only
- ▶ **Both I & II** (P# 181)
- ▶ None of I & II

67.

The key idea behind _____ is to take a big task and break it into subtasks that can be processed concurrently on a stream of data inputs in multiple, overlapping stages of execution.

- ▶ **Pipeline Parallelism** page no 214
- ▶ Overlapped Parallelism
- ▶ Massive Parallelism
- ▶ Distributed Parallelism

68.

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

- ▶ **Skew in Partition** (P # 218)
- ▶ Pipeline Distribution
- ▶ Distributed Distribution
- ▶ Uncontrolled Distribution

AL-JUNAID INSTITUTE GROUP

69.

The goal of ideal parallel execution is to completely parallelize those parts of a computation that are not constrained by data dependencies. The smaller the portion of the program that must be executed _____, the greater the scalability of the computation.

▶ None of these

▶ **Sequentially** page no 204

▶ In Parallel

70.

If 'M' rows from table-A match the conditions in the query then table-B is accessed 'M' times. Suppose table-B has an index on the join column. If 'a' I/Os are required to read the data block for each scan and 'b' I/Os for each data block then the total cost of accessing table-B is _____ logical I/Os approximately.

▶ **$(a + b)M$**

▶ $(a - b)M$

▶ $(a + b + M)$

▶ $(a * b * M)$

71.

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** page no 249

▶ Non-Exploratory

▶ Computer Science

72. 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** page no 254

▶ OLAP

▶ DSS

▶ DWH

73.

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

▶ **Clustering** [page no 264]

▶ Aggregation

▶ Segmentation

▶ Partitioning

74.

To measure or quantify the similarity or dissimilarity, different techniques are available. Which of the following option represent the name of available techniques?

▶ Pearson correlation is the only technique

▶ Euclidean distance is the only technique

▶ **Both Pearson correlation and Euclidean distance** [page no 270]

AL-JUNAID INSTITUTE GROUP

- ▶ None of these

75.

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

- ▶ **One-way Clustering** (P# 271)
- ▶ Bi-clustering
- ▶ Pearson correlation
- ▶ Euclidean distance

76.

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

- ▶ Designing
- ▶ **Development**
- ▶ Analysis
- ▶ Implementation

77.

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

- ▶ Tools
- ▶ **Industry** (P# 320)
- ▶ Software
- ▶ None of these

78.

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

- ▶ Increasing
- ▶ **Decreasing** (P# 215)
- ▶ Maintaining
- ▶ None of these

79.

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

- ▶ **Silver Bullet** (pg#375)
- ▶ Golden Bullet
- ▶ Suitable Hardware
- ▶ Compatible Product

80.

Focusing on data warehouse delivery only often end up _____.

- ▶ **Rebuilding** (pg# 315)
- ▶ Success
- ▶ Good Stable Product
- ▶ None of these

82.

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

- ▶ **Cotton-growing** (p # 330)
- ▶ Rice-growing
- ▶ Weapon Producing

AL-JUNAID INSTITUTE GROUP

83.

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

▶ **Data profiling** (P# 439)

▶ Data Anomaly Detection

▶ Record Duplicate Detection

▶ None of these

84.

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** (p # 19)

▶ Two Direction

▶ None of these

85.

DSS queries do not involve a primary key

▶ **True** (p # 21)

▶ False

86.

_____ contributes 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** (P# 330)

▶ Missing Data

▶ Data Stored in Heterogeneous Sources

87.

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

▶ **OLE DB** (p # 373)

▶ OLAP

▶ OLTP

▶ Data Warehouse

88.

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** (p #373)

▶ Documentations

▶ Guideline

89.

AL-JUNAID INSTITUTE GROUP

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 (p # 419)
- ▶ Rolled back

90. 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 (p # 419)
- ▶ Creation of package
- ▶ Connection of package

91.

To judge effectiveness we perform data profiling twice.

- ▶ One before Extraction and the other after Extraction
- ▶ One before Transformation and the other after Transformation [p # 441]
- ▶ One before Loading and the other after Loading

92.

During the development phase we should follow standards for:

- Naming Conventions
- Calculation
- Libraries
- All of the given options (Page 307)

93.

As per kimball's approach the second phase of DWH life cycle is:

- Project Planning
- Requirement definition (Page 299)
- Requirement verification
- Requirement validation

94.

Bill Inmon argues that requirements are well understood only after

- They are documented
- They extracted and verified

AL-JUNAID INSTITUTE GROUP

➤ **Data warehouse is populated (Page 285)**

➤ All schemes are defined

95.

Waterfall model is appropriate when

- When budget is low
- When the deadline is strict
- When resources are limited

➤ **Requirements are clearly defined (Page 284)**

96.

A bottom up implementation approach is useful when

- **Technology is matured and well understands (page 283)**
- Organization cannot implement latest technologies
- The business objectives are very much clear
- All requirements are well documented

97.

Technical architecture design specifies the:

- **Project schedule (Page 300)**
- Minimum project completion time
- Required Components
- Umbrella activities

98.

_____ says that as far as company goals are concerned, the entire company pursues in the same direction

➤ **Wester man (Page 285)**

- Bill Inmon
- Kimball
- Bohnlein

99.

As per Bill Inmon, a data warehouse, in contrast with classical applications is:

➤ **Data driven (Page 285)**

- Resource driven
- Requirement driven
- Time Sensitive

AL-JUNAID INSTITUTE GROUP

100.

The _____ phase of fundamental data warehouse life cycle model includes data warehouse scheduled maintenance

- Deployment
- **Operation (Page 288)**
- Enhancement
- Maintenance

101.

Following the coding standards during development helps to:

- **Minimize the future rework (Page 307)**
- Verify the requirements
- To refine the project scope

102.

As per Kimball, _____ is the main operational process

- Requirement extraction
- Global design
- **Business process (Page 285)**
- Scheme design

102.

In _____ phase of a fundamental data warehouse life cycle model, a working model of data warehouse is deployed for a selective set of users

- Design
- **Prototype (Page 287)**
- Deployment
- Operation

103.

Waterfall is a/an _____ model

- Iterative
- **Simple linear sequential (Page 284)**
- Object oriented
- Rapid development

104.

One of the drawbacks of waterfall model is that:

AL-JUNAID INSTITUTE GROUP

➤ **Customers can not review the product during development** (Page 284)

- It does not work when the resources are limited
- It does not define the project timeline/schedule
- All of the given options

105. Spiral model is _____

- Sequence of waterfall model
- Risk oriented model
- An iterative model
- **All of the given** (Page 284)

106.

_____ refers to the overall process of discovering useful knowledge from data and data mining refers to a particular step in this process.

- Statistics
- **Knowledge discovery in database** (Page 249)
- Clustering
- Information cleansing

107. Data mining is all about

- Knowledge discovery in database
- Finding hidden patterns in data
- Finding hidden relations in data
- **All of the given options** (Page 249)

108. In context of clustering, the term "distance" means

- **Similarity/dissimilarity of records** (Page 272)
- The difference between the primary keys of two records
- The relation of a record with corresponding record in child table
- None of the given options

109.

We should follow a proper _____ cycle to implement a change even if it is smaller one.

- **Development -> QA -> Production** (Page 314)
- Production -> QA -> Development
- Development -> Production -> QA
- Production->Development -> QA

110.

For smooth DWH project implementation, one of the recommendations is to have:

- **Full time project manager assigned to the project** (Page 318)

AL-JUNAID INSTITUTE GROUP

- QA team ready for testing the code
- Design of complete project before implementation
- Components of complete project before implementation

111.

Normally it is recommended to have

- Different servers for development and production environment
- Same server for development and production environment
- **Interference while having different database environments on a single server (Page 313)**
- All of the given options

112.

Identify the TRUE statement:

- **Clustering is unsupervised learning and classification is supervised learning (Page 270)**
- Clustering is supervised learning and classification is unsupervised learning
- Both clustering and classification are unsupervised learning
- Both clustering and classification are supervised learning

113. Mining multi-dimensional databases allow users to:

- Categorize the data
- **All of the given options (250)**
- Analyze the data
- Summarize the data

114.

Technical architecture design supports the communication about technical requirements: I) within the team II) Upward to management III) Outward to vendors

- only
- Only
- Only
- **(I), (II) and (III) (Page 300)**

115.

Improper documentation results the problem(s) like:

- Maintenance issue
- **New developers unable to configure already existing code (Page 314)**
- Lot of time required for enhancing the code
- All of the given options

AL-JUNAID INSTITUTE GROUP

116.

In Four Cell Quadrant Technique, The quadrant's vertical axis refers to:

- Scope
- Feasibility
- Resources available
- **The potential impact or value to the business (Page 297)**

117.

Which of the following activity executes parallel with all other activities in Kimball's DWH development approach?

- Requirement elicitation
- Project Planning
- **Project management (Page 289)**
- Deployment

118.

"What means what". The phrase refers to:

- **Meta data (Page 338)**
- External data
- Transformed data
- Internal representations

119.

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 (Page 316)**
- Isolating IT support people from business users
- Training the users with dummy data and considering it success

120.

In _____ phase of kimball's approach, we identify the components needed now and in future.

- Requirement definition
- **Architectural design (Page 300)**
- Product development
- Analytical application development

121.

Which of the following is NOT one of the advantages of changed data capture (CDC) technique?

AL-JUNAID INSTITUTE GROUP

- 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

122.

_____ technique requires a separate column to specify the time and date when the last modification was occurred.

- Check Marks
- **Time Stamps (Page 150)**
- Just-In-Time
- Real Time extraction

123.

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

- **Syntactically dirty data (Page 160)**
- Semantically dirty data
- Coverage anomaly
- Extraction issue

124.

In Extract, Load, Transform (ELT) process, data transformation _____

- Takes place on the data warehouse server
- **Takes place on a separate transformation server (Page 147)**
- Depends on the nature of the source database
- Does not take place

125.

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

126.

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

- Conversion

AL-JUNAID INSTITUTE GROUP

- Summarization
- Enrichment
- **Full Data Refresh (Page 135)**

127.

Robotic libraries are needed for _____

- Cubes
- Data marts
- **Data warehouse (Page 131)**
- Aggregates

128.

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

- Aggregates don't change in real time
- **Transformation of extracted data is difficult (Page 149)**
- Identifying the recently modified data may be difficult
- Source systems may not support extraction of changed aggregates

129.

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 (Page 143)**
- Prioritization
- Element Selection
- Measurement Selection

130.

In context of Change Data Capture (CDC), sometimes a _____ object can be used to store the recently modified data:

- Buffer table
- **Change table (Page 149)**
- Checkmark table
- Change control table

131.

In data mining, initially you _____ what you are looking for.

- Know
- **Don't Know (Page 250)**

AL-JUNAID INSTITUTE GROUP

- May or may not know
- None of the given options

132.

In context of the most fundamental data warehouse life cycle model, which of the following is NOT one of the data warehouse design activities?

- End-user interviews and re-interviews
- Source system cataloguing
- Definition of key performance indicators
- **System vision development (Page 287)**

133.

A top down implementation approach is useful when

- **Technology is matured and well understands (Page 283)**
- Organization cannot implement latest technologies
- The business objectives are very much clear
- All requirements are well documented

134.

Which of the following is NOT one of the methodologies for Data Warehouse project development?

- Goal Driven
- Data Driven
- User Driven
- **System Driven (Page 283)**

135. In context of data mining definition, the term "value" means

- The primary key of table
- Index location of the record
- **Importance of hidden patterns discovered (Page 250)**
- Numerical or string measure assigned to an attribute

136

Data mining is all about

- Knowledge discovery in database
- Finding hidden patterns in data
- Finding hidden relations in data
- **All of the given options (Page 249)**

137.

AL-JUNAID INSTITUTE GROUP

In context of clustering, the term "distance" means

- **Similarity/dissimilarity of records (Page 272)**
- The difference between the primary keys of two records
- The relation of a record with corresponding record in child table
- None of the given options

138.

We should follow a proper _____ cycle to implement a change even if it is smaller one.

- **Development -> QA -> Production (Page 314)**
- Production -> QA -> Development
- Development -> Production -> QA
- Production->Development -> QA

139.

For smooth DWH project implementation, one of the recommendations is to have:

- **Full time project manager assigned to the project (Page 318)**
- QA team ready for testing the code
- Design of complete project before implementation
- Components of complete project before implementation

140.

Normally it is recommended to have

- Different servers for development and production environment
- Same server for development and production environment
- **Interference while having different database environments on a single server (Page 313)**
- All of the given options

141.

Identify the TRUE statement:

- **Clustering is unsupervised learning and classification is supervised learning (Page 270)**
- Clustering is supervised learning and classification is unsupervised learning
- Both clustering and classification are unsupervised learning
- Both clustering and classification are supervised learning

142.

Mining multi-dimensional databases allow users to:

AL-JUNAID INSTITUTE GROUP

- Categorize the data
- Analyze the data
- Summarize the data
- **All of the given options (Page 250)**

143.

Technical architecture design supports the communication about technical requirements: I) within the team II) Upward to management III) Outward to vendors

- only
- Only
- Only
- **(I), (II) and (III) (Page 300)**

144.

Improper documentation results the problem(s) like:

- Maintenance issue
- **New developers unable to configure already existing code (Page 314)**
- Lot of time required for enhancing the code
- All of the given options

145.

In Four Cell Quadrant Technique, The quadrant's vertical axis refers to:

- Scope
- Feasibility
- Resources available
- **The potential impact or value to the business (Page 297)**

146.

Which of the following activity executes parallel with all other activities in Kimball's DWH development approach?

- Requirement elicitation
- Project Planning
- **Project management (Page 289)**
- Deployment

147.

"What means what". The phrase refers to:

- **Meta data (Page 338)**
- External data
- Transformed data

AL-JUNAID INSTITUTE GROUP

- Internal representations

148.

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 (Page 316)**
- Isolating IT support people from business users
- Training the users with dummy data and considering it success

149.

In _____ phase of kimballs approach, we identify the components needed now and in future.

- Requirement definition
- **Architectural design (Page 300)**
- Product development
- Analytical application development

150.

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

151.

_____ technique requires a separate column to specify the time and date when the last modification was occurred.

- Check Marks
- **Time Stamps (Page 150)**
- Just-In-Time

- Real Time extraction

152.

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

- **Syntactically dirty data (Page 160)**
- Semantically dirty data

AL-JUNAID INSTITUTE GROUP

- Coverage anomaly
- Extraction issue

153.

In Extract, Load, Transform (ELT) process, data transformation _____

- Takes place on the data warehouse server
- **Takes place on a separate transformation server (Page 147)**
- Depends on the nature of the source database
- Does not take place

154.

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

155.

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

- Conversion
- Summarization
- Enrichment
- **Full Data Refresh (Page 135)**

156.

Robotic libraries are needed for _____

- Cubes
- Data marts
- **Data warehouse (Page 131)**
- Aggregates

157.

During the development phase we should follow standards for:

AL-JUNAID INSTITUTE GROUP

- Naming Conventions
- Calculation
- Libraries
- All of the given options (Page 307)

158.

As per kimball's approach the second phase of DWH life cycle is:

- Project Planning
- Requirement definition (Page 299)
- Requirement verification
- Requirement validation

159.

Bill Inmon argues that requirements are well understood only after

- They are documented
- They extracted and verified
- Data warehouse is populated (Page 285)
- All schemes are defined

160.

Waterfall model is appropriate when

- When budget is low
- When the deadline is strict
- When resources are limited
- Requirements are clearly defined (Page 284)

161.

A bottom up implementation approach is useful when

- Technology is matured and well understands (page 283)
- Organization cannot implement latest technologies
- The business objectives are very much clear
- All requirements are well documented

162.

Technical architecture design specifies the:

- Project schedule (Page 300)
- Minimum project completion time
- Required Components

- Umbrella activities

163.

AL-JUNAID INSTITUTE GROUP

_____ says that as far as company goals are concerned, the entire company pursues in the same direction

- Westerman (Page 285)
- Bill inmon
- Kimball
- Bohnlein

164.

As per Bill Inmon, a data warehouse, in contrast with classical applications is:

- Data driven (Page 285)
- Resource driven
- Requirement driven
- Time Sensitive

165.

The _____ phase of fundamental data warehouse life cycle model includes data warehouse scheduled maintenance

- Deployment
- Operation (Page 288)
- Enhancement
- Maintenance

166.

Following the coding standards during development helps to:

- Minimize the future rework (Page 307)
- Verify the requirements
- To refine the project scope
- To establish the test cases

167.

As per Kimball, _____ is the main operational process

- Requirement extraction
- Global design
- Business process (Page 285)
- Scheme design

168..

In _____ phase of a fundamental data warehouse life cycle model, a working model of data warehouse is deployed for a selective set of users

- Design
- Prototype (Page 287)

AL-JUNAID INSTITUTE GROUP

- Deployment
- Operation

169. Waterfall is a/an _____ model

- Iterative
- Simple linear sequential (Page 284)
- Object oriented
- Rapid development

170. One of the drawbacks of waterfall model is that:

- Customers can not review the product during development (Page 284)
- It does not work when the resources are limited
- It does not define the project timeline/schedule
- All of the given options

171.

Spiral model is _____

- Sequence of waterfall model
- Risk oriented model
- An iterative model
- All of the given (Page 284)

172.

_____ refers to the overall process of discovering useful knowledge from data and data mining refers to a particular step in this process.

- Statistics
- Knowledge discovery in database (Page 249)
- Clustering
- Information cleansing

173.

In data mining, initially you _____ what you are looking for.

- Know
- Don't Know (Page 250)
- May or may not know
- None of the given options

174.

In context of the most fundamental data warehouse life cycle model, which of the following is NOT one of the data warehouse design activities?

- End-user interviews and re-interviews
- Source system cataloguing

AL-JUNAID INSTITUTE GROUP

- Definition of key performance indicators

- System vision development (Page 287)

175.

A top down implementation approach is useful when

- Technology is matured and well understands (Page 283)
- Organization cannot implement latest technologies
- The business objectives are very much clear
- All requirements are well documented

176.

Which of the following is NOT one of the methodologies for Data Warehouse project development?

- Goal Driven
- Data Driven
- User Driven
- System Driven (Page 283)

177.

In context of data mining definition, the term "value" means

- The primary key of table
- Index location of the record
- Importance of hidden patterns discovered (Page 250)
- Numerical of string measure assigned to an attribute

178.

Data mining is all about

- Knowledge discovery in database
- Finding hidden patterns in data
- Finding hidden relations in data
- All of the given options (Page 249)

179.

In context of clustering, the term "distance" means

- Similarity/dissimilarity of records (Page 272)
- The difference between the primary keys of two records
- The relation of a record with corresponding record in child table
- None of the given options

180.

We should follow a proper _____ cycle to implement a change even if it is smaller one.

AL-JUNAID INSTITUTE GROUP

➤ Development -> QA -> Production (Page 314)

- Production -> QA -> Development
- Development -> Production -> QA
- Production->Development -> QA

181.

For smooth DWH project implementation, one of the recommendations is to have:

➤ Full time project manager assigned to the project (Page 318)

- QA team ready for testing the code
- Design of complete project before implementation
- Components of complete project before implementation

182.

Normally it is recommended to have

- Different servers for development and production environment
- Same server for development and production environment
- Interference while having different database environments on a single server (Page 313)
- All of the given options

183.

Identify the TRUE statement:

- Clustering is unsupervised learning and classification is supervised learning (Page 270)
- Clustering is supervised learning and classification is unsupervised learning
- Both clustering and classification are unsupervised learning
- Both clustering and classification are supervised learning

184.

Mining multi-dimensional databases allow users to:

- Categorize the data
- Analyze the data
- Summarize the data
- All of the given options (Page 250)

185.

Technical architecture design supports the communication about technical requirements: I) within the team II) Upward to management III) Outward to vendors

- only
- Only
- Only

AL-JUNAID INSTITUTE GROUP

➤ (I), (II) and (III)

(Page 300)

186.

Improper documentation results the problem(s) like:

➤ Maintenance issue

➤ New developers unable to configure already existing code (Page 314)

➤ Lot of time required for enhancing the code

➤ All of the given options

187.

In Four Cell Quadrant Technique, The quadrant's vertical axis refers to:

➤ Scope

➤ Feasibility

➤ Resources available

➤ The potential impact or value to the business (Page 297)

188.

Which of the following activity executes parallel with all other activities in Kimball's DWH development approach?

➤ Requirement elicitation

➤ Project Planning

➤ Project management (Page 289)

➤ Deployment

189.

"What means what". The phrase refers to:

➤ Meta data (Page 338)

➤ External data

➤ Transformed data

➤ Internal representations

190.

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 (Page 316)

➤ Isolating IT support people from business users

➤ Training the users with dummy data and considering it success

191.

In _____ phase of kimballs approach, we identify the components needed now and in future.

AL-JUNAID INSTITUTE GROUP

- Requirement definition
- **Architectural design (Page 300)**
- Product development
- Analytical application development

192.

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

193.

_____ technique requires a separate column to specify the time and date when the last modification was occurred.

- Check Marks
- **Time Stamps (Page 150)**
- Just-In-Time
- Real Time extraction

194.

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

- **Syntactically dirty data (Page 160)**
- Semantically dirty data
- Coverage anomaly
- Extraction issue

195.

In Extract, Load, Transform (ELT) process, data transformation _____

- Takes place on the data warehouse server
- **Takes place on a separate transformation server (Page 147)**
- Depends on the nature of the source database
- Does not take place

196.

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

- Conversion
- Summarization

AL-JUNAID INSTITUTE GROUP

- Enrichment
- Full Data Refresh (Page 135)

197.

Robotic libraries are needed for _____

- Cubes
- Data marts
- Data warehouse (Page 131)
- Aggregates

198.

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

- Aggregates don't change in real time
- Transformation of extracted data is difficult (Page 149)
- Identifying the recently modified data may be difficult
- Source systems may not support extraction of changed aggregate

199.

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 (Page 143)
- Prioritization
- Element Selection
- Measurement Selection

200.

In context of Change Data Capture (CDC), sometimes a ____ object can be used to store the recently modified data:

- Buffer table
- Change table (Page 149)
- Checkmark table
- Change control table

201.

Which of the following is NOT one of the examples of dynamic attributes?

- Daily Sale
- Date of Birth (confirm not) (Page 342)
- Air pressure
- None of the given options

202.

In context of requirement definition phase of Kimball's DWH development approach, _____ is positioned as a findings review and prioritization meeting.

AL-JUNAID INSTITUTE GROUP

- System analysis
- Scope definition
- Requirement configuration
- Requirements wrap-up presentation (Page 297)

203.

Which of the following activity/activities is/are part of project planning phase in Kimball's DWH development approach?

- Obtain resources
- Establish the preliminary scope and justification
- Assess organization's readiness for a data warehouse initiative
- All of the given options (Page 290)

204.

Goal driven approach of data warehouse development was result of _____ work

- Bill Inmon
- Ralph Kimball
- Böhnlein and Ulbrich-vom (Page 285)
- Westerman

205.

In contrast to data mining, statistics is _____ driven.

- Assumption
- Knowledge
- human (Page 255)
- Database

206.

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

- Linear
- Quadratic
- Exponential (Page 15)
- Logarithmic

207.

In context of requirement definition phase in Kimball's DWH development approach, activities like debriefing, documentation, and prioritization are considered to be the part of:

- Requirement preplanning
- Business requirements collection
- Post collection (Page 294)
- None of the given options

AL-JUNAID INSTITUTE GROUP

208.

Users do not care, how advance the front end of your DWH is, what they care is that:

- Tables should be properly de-normalized
- Proper partitioning technique should be used
- At least star or snow flake schema should be implementing
- They should get information in timely manner (Page 314)

209.

Which of the following is NOT one of the activities of “Maintenance and Growth” phase in Kimball’s DWH development approach?

- Education
- Technical education
- Program Support
- Interface Deployment (Page 309)

210.

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

- Employee Name (Page 342)
- Employee Date of Birth
- Employee Blood Group
- All of the given options

211.

Three parallel tracks in kimbal’s DWH development approach converge at _____ Phase

- Project planning
- Business requirement definition
- Deployment (Page 308)
- Maintenance and growth

212.

Which of the following is not a technique of "De-Normalization" ?

- Pre-joining
- Splitting Tables
- Adding Redundant Columns
- ER Modeling (Page 52)

213.

Which one of the following is not a technique of "Change Data Capture” in currently used Modem Source Systems?

- Timestamps
- Partitioning
- Triggers

AL-JUNAID INSTITUTE GROUP

- Dimensional Modeling (Page 150)

214.

Normally the term "DWH face to the business user" refers to:

- Lifecycles technology track
- Lifecycle Data track
- Lifecycle Analytical Applications track (Page 306)
- Lifecycle Maintenance track

215.

An effective user education program includes, among other, the following guideline(s):

- Understand the audience, don't overwhelm
- Train after delivery of the data and analytic applications
- Postpone education, if DWH not ready
- All of the given options (Page 308)

216.

In context of analytical applications track, the application development activity can begin after

- Database Design is complete (Page 307)
- A subset of historical data has been loaded
- The data access tools and metadata are installed
- All of the given options

217.

During the development phase we should follow standards for:

- Naming Conventions
- Libraries
- Calculation
- All of the given options (page 307)

218.

As per Kimball's approach the second phase of DWH life cycle is:

- Project Planning
- Requirement definition (page 299)
- Requirement verification
- Requirement validation

219.

Bill Inmon argues that requirements are well understood only after

- They are documented
- They extracted and verified

AL-JUNAID INSTITUTE GROUP

- Data warehouse is populated (page 285)
- All schemes are defined

220.

Waterfall model is appropriate when

- When budget is low
- When the deadlines is strict
- When resources are limited
- Requirements are clearly defined (page 284)

221.

A bottom up implementation approach is useful when

- Technology is matured and well understands (page 283)
- Organization cannot implement latest technologies
- The business objectives are very much clear
- All requirements are well documented

222.

Technical architecture design specifies the:

- Project schedule (page 300)
- Minimum project completion time
- Required components
- Umbrella activities

223.

----- says that as far as company goals are connected, the entire company pursues in the same direction

- Westerman (page 285)
- Bill inmon
- Kimball
- Bohnlein

224.

As per Bill Inmon, a data warehouse, in contrast with classical applications is:

- Data driven (page 285)
- Resource driven
- Requirement driven
- Time sensitive

225.

AL-JUNAID INSTITUTE GROUP

The.....phase of fundamental data warehouse life cycle model includes data warehouse scheduled maintenance

- Deployment
- Operation (page 288)
- Enhancement
- Maintenance

226.

Following the coding standards during development helps to:

- Minimize the future work (page 307)
- Verify the requirements
- To refine the project scope
- To establish the test cases

227.

As per Kimball,is the main operational process.

- Requirements extraction
- Global design
- Business process (page 285)
- Scheme design

228.

In phase of fundamental data warehouse life cycle model, a working model of data warehouse is deployed for a selective set of users.

- Design
- Prototype (page 287)
- Deployment
- Operation

229.

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 (page 143)
- Prioritization
- Element Selection
- Measurement Selection

230.

Goal driven approach of data warehouse development was result of work.

- Bill Inmon
- Ralph Kimball
- Bohnlein and ulbrich-vom (page 285)

AL-JUNAID INSTITUTE GROUP

- Westerman

231.

In contrast to data mining, statistics is _____ driven.

- Assumption
- Knowledge
- **Human** (page 255)
- Database

232.

Which of the following is example of static attributes?

- **Employee Name** (page 342)
- Employee date of birth
- Employee blood group
- All of the given

233.

In context of analytical applications track, the application development activity can design after

- **Database design is complete** (page 307)
- A subset of historical data has been loaded
- The data access tools and metadata are installed
- All of the given options

234. Bill Inmon argues that requirements are well understood only after

- They are documented
- They extracted and verified
- Data warehouse is populated** page 285
- All the schemas are defined

235. Spiral model is _____

- Sequence of waterfall model** page 284
- Risk oriented model
- An iterative model
- All of the given options

236.

Waterfall is a/an _____ model

- Iterative
- Simple linear sequential** page 284
- Object oriented
- Rapid development

237. An effective user education program includes, among others, the

AL-JUNAID INSTITUTE GROUP

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**
238. 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**
239. As per Bill Inmon, a data warehouse, in contrast with classical applications is:
- a. Data driven page 285**
 - b. Resource driven
 - c. Requirement driven
 - d. Time sensitive
240. 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
241. In context of requirement definition phase of Kimball's DWH development approach, ___ 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**
242. 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**

AL-JUNAID INSTITUTE GROUP

243. 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

244. 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

245. 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

246. Which of the following is NOT one of the possible pitfalls in DWH Life

Cycle & 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

247. 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

248. Which of the following is/are included in the list of Top-7 key steps for smooth DWH implementation?

AL-JUNAID INSTITUTE GROUP

- 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

249. 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

250. 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

251. In context of requirement definition phase in Kimball's DWH development approach, 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



AL-JUNAID INSTITUTE GROUP

252. Which of the following is NOT one of the three parallel tracks in Kimball's approach?

- a. Lifecycle technology track
- b. Lifecycle data track
- c. Lifecycle analytical applications track

d. Lifecycle maintenance track page 299

253. 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

254. 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

255. 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

256. 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

257. As per Kimball, ___ is the main operational process

- a. Requirement extraction
- b. Goal design

c. Business process page 285

d. Schema design

258. Which of the following is NOT one of the top-10 mistakes that should be avoided during DWH development?

- a. Not interacting directly with end user

AL-JUNAID INSTITUTE GROUP

- 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



AL-JUNAID INSTITUTE GROUP

259. The _____ phase of fundamental data warehouse life cycle model includes data warehouse daily maintenance activities
- a. Deployment
 - b. Operation page 288**
 - c. Enhancement
 - d. Maintenance
260. In context of the most fundamental data warehouse life cycle model, which of the following is NOT one of the data warehouse design activities?
- a. End-user interviews and re-interviews
 - b. Source system cataloging
 - c. Definition of key performance indicators
 - d. System vision development page 287**
261. 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**
262. In context of analytical applications track, the application development activity 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**
263. 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**
264. Users do not care, how advance the front end of your DWH is, what they care 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 they**

AL-JUNAID INSTITUTE GROUP

want

page 314



AL-JUNAID INSTITUTE GROUP

265. Which of the following activity executes parallel with all other activities in Kimball's DWH development approach?

- a. Requirement elicitation
- b. Project planning
- c. Project management**
- d. Deployment

266. 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**

267. Which of the following is NOT one of the activities of "Maintenance and Growth" phase in Kimball's DWH development approach?

- a. Education
- b. Technical Education
- c. Program Support
- d. Interface Deployment** **page 309**

268. In _____ phase of Kimball's approach, we identify the components needed now and in future.

- a. Requirement definition
- b. Architectural design** **page 300**
- c. Product development
- d. Analytical application development

269. Implementation of a data warehouse requires _____ activities

- a. Highly integrated
- b. Loosely integrated
- c. Tightly decoupled
- d. None of the given** **page 289**

270. Which of the following activity/activities is/are part of project planning phase in Kimball's DWH development approach?

- a. Obtain resources
- b. Establish the preliminary scope and justification
- c. Assess organization's readiness for a data warehouse initiative
- d. All of the given options** **page 290**

AL-JUNAID INSTITUTE GROUP

271. _____ says that as far as company goals are concerned, the entire company pursues in the same direction
- a. Westerman page 285**
 - b. Bill Inmon
 - c. Kimball
 - d. Bohnlein
272. To identify the _____ required we need to perform data profiling
- ▶ Degree of Transformation (Page 437)**
 - ▶ Complexity
 - ▶ Cost
 - ▶ Time\
273. 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 of these
 - ▶ Internal
 - ▶ External (Page 21)**
274. Ad-hoc access means to run such queries which are known already.
- ▶ True
 - ▶ False (Page 18)**
275. one _____ in agriculture extension is that pest population beyond which the benefit of spraying outweighs its cost.
- ▶ Profit Threshold Level
 - ▶ Economic Threshold Level (Page 332)**
 - ▶ Medicine Threshold Level
 - ▶ None of these
276. People that design and build the data warehouse must be capable of working across the organization at all levels
- ▶ True (Page 320)**
 - ▶ False
277. The _____ is only a small part in realizing the true business value buried within the mountain of data collected and stored within organizations business systems and operational databases.
- ▶ Independence on technology
 - ▶ Dependence on technology (Page 320)**
 - ▶ None of these
278. one 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

AL-JUNAID INSTITUTE GROUP

- ▶ Any Direction (Page 19)
 - ▶ Two Direction
 - ▶ None of these
279. DSS queries do not involve a primary key
- ▶ True (Page 21)
 - ▶ False
280. 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 (Page 373)
 - ▶ Documentations
 - ▶ Guidelines
281. The need to synchronize data upon update is called
- ▶ Data Manipulation
 - ▶ Data Replication
 - ▶ Data Coherency (Page 12)
 - ▶ Data Imitation
282. Taken jointly, the extract programs or naturally evolving systems formed a spider web, also known as
- ▶ Distributed Systems Architecture
 - ▶ Legacy Systems Architecture (Page 14)
 - ▶ Online Systems Architecture
 - ▶ Intranet Systems Architecture
283. Node of a B-Tree is stored in memory block and traversing a B-Tree involves _____ page faults.
- ▶ $O(n)$
 - ▶ $O(n^2)$
 - ▶ $O(n \lg n)$
 - ▶ $O(\lg n)$ (Page 22)
284. 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 benefit at both query time and update time.
 - ▶ Redundant data is a performance liability at both query time and update time.
 - ▶ Redundant data is a performance benefit at query time, but is a performance liability at update time. (Page 51)
285. one As apposed to the out come of classification, estimation deal with _____ valued outcome.
- ▶ Discrete
 - ▶ Isolated
 - ▶ Continuous (Page 260)

AL-JUNAID INSTITUTE GROUP

- ▶ Distinct
286. The goal of _____ is to look at as few block as possible to find the matching records.
- ▶ Indexing (Page 222)
 - ▶ Partitioning
 - ▶ Joining
 - ▶ none of these
287. The technique that is used to perform these feats in data mining modeling, and this act of model building is something that people have been doing for long time, certainly before the _____ of computers or data mining technology.
- ▶ Access Advent
 - ▶ Ascent Avowal
288. De-Normalization normally speeds up _____
- ▶ Data Retrieval (Page 51)
 - ▶ Data Modification
 - ▶ Development Cycle
 - ▶ Data Replication
289. one In horizontal splitting, we split a relation into multiple tables on the basis of _____
- ▶ Common Column Values (Page 54)
 - ▶ Common Row Values
 - ▶ Different Index Values
 - ▶ Value resulted by ad-hoc query
290. one 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 (Page 16)
 - ▶ Decision Maker's Perspective
 - ▶ Employee's Perspective
291. The automated, prospective analyses offered by data mining move beyond the analysis of past events provided by respective tools typical of _____.
- ▶ OLTP
 - ▶ OLAP
 - ▶ Decision Support systems
 - ▶ None of these
292. A data warehouse may include _____
- ▶ Legacy systems (Page 135)
 - ▶ Only internal data sources
 - ▶ Privacy restrictions
 - ▶ Small data mart
293. De-Normalization normally speeds up _____
- ▶ Data Retrieval (Page 51)
 - ▶ Data Modification

AL-JUNAID INSTITUTE GROUP

- ▶ Development Cycle
 - ▶ Data Replication
294. In horizontal splitting, we split a relation into multiple tables on the basis of
- ▶ Common Column Values (Page 54)
 - ▶ Common Row Values
 - ▶ Different Index Values
 - ▶ Value resulted by ad-hoc query
295. one Multidimensional databases typically use proprietary _____ format to store pre-summarized cube structures.
- ▶ File (Page 79)
 - ▶ Application
 - ▶ Aggregate
 - ▶ Database
296. A dense index, if fits into memory, costs only _____ disk I/O access to locate a record by given key.
- ▶ One (Page 223)
 - ▶ Two
 - ▶ $\lg(n)$
 - ▶ n
297. All data is _____ of something real. I An Abstraction II A Representation
Which of the following option is true?
- ▶ I Only (Page 180)
 - ▶ II Only
 - ▶ Both I & II
 - ▶ None of I & II
298. The key idea behind _____ is to take a big task and break it into subtasks that can be processed concurrently on a stream of data inputs in multiple, overlapping stages of execution.
- ▶ Pipeline Parallelism (Page 214)
 - ▶ Overlapped Parallelism
 - ▶ Massive Parallelism
 - ▶ Distributed Parallelism
299. Non uniform distribution, when the data is distributed across the processors, is called _____.
- ▶ Skew in Partition (Page 218)
 - ▶ Pipeline Distribution
 - ▶ Distributed Distribution
 - ▶ Uncontrolled Distribution
300. one The goal of ideal parallel execution is to completely parallelize those parts of a computation that are not constrained by data dependencies. The smaller the portion of the program that must be executed _____, the greater the scalability of the computation.

AL-JUNAID INSTITUTE GROUP

- ▶ None of these
 - ▶ **Sequentially (Page 204)**
 - ▶ In Parallel
 - ▶ Distributed
301. If 'M' rows from table-A match the conditions in the query then table-B is accessed 'M' times. Suppose table-B has an index on the join column. If 'a' I/Os are required to read the data block for each scan and 'b' I/Os for each data block then the total cost of accessing table-B is _____ logical I/Os approximately.
- ▶ **$(a + b)M$**
 - ▶ $(a - b)M$
 - ▶ $(a + b + M)$
 - ▶ $(a * b * M) / 7$
302. 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 (Page 249)**
 - ▶ Non-Exploratory
 - ▶ Computer Science
303. 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 (Page 254)**
 - ▶ OLAP
 - ▶ DSS
 - ▶ DWH
304. _____ is the technique in which existing heterogeneous segments are reshuffled, relocated into homogeneous segments.
- ▶ **Clustering (Page 264)**
 - ▶ Aggregation
 - ▶ Segmentation
 - ▶ Partitioning
305. To measure or quantify the similarity or dissimilarity, different techniques are available. Which of the following option represent the name of available techniques?
- ▶ Pearson correlation is the only technique
 - ▶ Euclidean distance is the only technique
 - ▶ **Both Pearson correlation and Euclidean distance (Page 270)**
 - ▶ None of these
306. 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

AL-JUNAID INSTITUTE GROUP

- ▶ Euclidean distance
307. In DWH project, it is assured that _____ environment is similar to the production environment
- ▶ Designing
 - ▶ **Development (Page 314)**
 - ▶ Analysis
 - ▶ Implementation
308. For a DWH project, the key requirements are _____ and product experience.
- ▶ Tools
 - ▶ **Industry (Page 320)**
 - ▶ Software
 - ▶ None of these
309. Pipeline parallelism focuses on increasing throughput of task execution, NOT on _____ sub-task execution time.
- ▶ Increasing
 - ▶ **Decreasing (Page 215)**
 - ▶ Maintaining
 - ▶ None of these
310. Many data warehouse project teams waste enormous amounts of time searching in vain for a _____.
- ▶ **Silver Bullet (Page 315)**
 - ▶ Golden Bullet
 - ▶ Suitable Hardware
 - ▶ Compatible Product
311. Focusing on data warehouse delivery only often ends up _____.
- ▶ **Rebuilding (Page 315)**
 - ▶ Success
 - ▶ Good Stable Product
 - ▶ None of these
312. One Pakistan is one of the five major _____ countries in the world.
- ▶ **Cotton-growing (Page 330)**
 - ▶ Rice-growing
 - ▶ Weapon Producing
313. _____ is a process which involves gathering of information about columns through execution of certain queries with intention to identify erroneous records.
- ▶ **Data profiling (Page 439)**
 - ▶ Data Anomaly Detection
 - ▶ Record Duplicate Detection
 - ▶ None of these

AL-JUNAID INSTITUTE GROUP

314. 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 (Page 19)
 - ▶ Two Direction
 - ▶ None of these
315. DSS queries do not involve a primary key
- ▶ True (Page 21)
 - ▶ False
316. _____ contributes 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 (Page 330)
 - ▶ Missing Data
 - ▶ Data Stored in Heterogeneous Sources
317. DTS allows us to connect through any data source or destination that is supported by _____
- ▶ OLE DB (Page 373)
 - ▶ OLAP
 - ▶ OLTP
 - ▶ Data Warehouse
318. 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 (Page 373) rep
 - ▶ Documentations
 - ▶ Guidelines
319. one 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 (Page 419)
 - ▶ Rolled back 10
320. 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 (Page 419)
 - ▶ Creation of package
 - ▶ Connection of package
321. To judge effectiveness we perform data profiling twice.
- ▶ One before Extraction and the other after Extraction
 - ▶ One before Transformation and the other after Transformation (Page 441)

AL-JUNAID INSTITUTE GROUP

- ▶ One before Loading and the other after Loading

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

- ▶ Doubles (Page 15)

- ▶ Triples
- ▶ Quartiles
- ▶ Remains same as previous year

323. Pre-computed _____ can solve performance problems

- ▶ Aggregates (Page 11)

- ▶ Facts
- ▶ Dimensions

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

- ▶ A Dense Index (Page 223)

- ▶ A Sparse Index
- ▶ An Inverted Index
- ▶ None of These

325. The degree of similarity between two records, often measured by a numerical value between _____, usually depends on application characteristics.

- ▶ 0 and 1 (Page 169)

- ▶ 0 and 10
- ▶ 0 and 100
- ▶ 0 and 99

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

- ▶ Two (Page 194)

- ▶ Three
- ▶ Four
- ▶ All

327. one NUMA stands for _____

- ▶ Non-uniform Memory Access (Page 206)

- ▶ Non-updateable Memory Architecture

AL-JUNAID INSTITUTE GROUP

▶ New Universal Memory Architecture

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

- ▶ Hash Join
- ▶ Inner Join
- ▶ Outer Join
- ▶ Sort-Merge Join

329. 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 (Page 243)
- ▶ None of these

330. Data mining derives its name from the similarities between searching for valuable business information in a large database, for example, finding linked products in gigabytes of store scanner data, and mining a mountain for a _____ of valuable ore.

- ▶ Furrow
- ▶ Streak
- ▶ Trough
- ▶ Vein

331. With data mining, the best way to accomplish this is by setting aside some of your data in a _____ to isolate it from the mining process; once the mining is complete, the results can be tested against the isolated data to confirm the model's validity.

- ▶ Cell
- ▶ Disk
- ▶ Folder
- ▶ Vault

332. The Kimball's iterative data warehouse development approach drew on decades of experience to develop the _____.

- ▶ Business Dimensional Lifecycle (Page 289)

AL-JUNAID INSTITUTE GROUP

▶ Data Warehouse Dimension

▶ Business Definition Lifecycle

▶ OLAP Dimension

333. We must try to find the one access tool that will handle all the needs of their users.

▶ True

▶ False (Page 315)

334. For a smooth DWH implementation we must be a technologist.

▶ True

▶ False (Page 319)

335. During the application specification activity, we also must give consideration to the organization of the applications.

▶ True (Page 307)

▶ False

336. 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

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

▶ Boll Worm (Page 333)

▶ Purple Worm

▶ Blue Worm

▶ Cotton Worm

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

▶ Decision maker (Page 18)

▶ Manager

▶ Database Administrator

▶ DWH Analyst

AL-JUNAID INSTITUTE GROUP

339. _____ breaks a table into multiple tables based upon common column values.

▶ Horizontal splitting (Page 54)

▶ Vertical splitting

340. 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 (Page 419) rep

▶ Rolled back

341. With data mining, the best way to accomplish this is by setting aside some of your data in a vault to isolate it from the mining process; once the mining is complete, the results can be tested against the isolated data to confirm the model's _____.

▶ Validity

▶ Security

▶ Integrity

▶ None of above

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

▶ Object

Container

▶ Subject

▶ Class (Page 259)

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

▶ OLAP

▶ DSS

▶ Data Mining

▶ OLTP (Page 206)

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

▶ Assumption

▶ Knowledge (Page 254)

AL-JUNAID INSTITUTE GROUP

▶ Human

▶ Database

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

▶ Discrete

▶ Isolated

▶ Continuous (Page 260)

▶ Distinct 15

346. The goal of ideal parallel execution is to completely parallelize those parts of a computation that are not constrained by data dependencies. The _____ the portion of the program that must be executed sequentially, the greater the scalability of the computation.

▶ Larger

▶ Smaller (Page 204)

▶ Unambiguous

▶ Superior

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

▶ Indexing (Page 222)

▶ Partitioning

▶ Joining

▶ None of above

348. In nested-loop join case, if there are 'M' rows in outer table and 'N' rows in inner table, time complexity is

▶ $O(M \log N)$

▶ $O(\log MN)$

▶ $O(MN)$

▶ $O(M + N)$

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

▶ Dense Index Page 223)

AL-JUNAID INSTITUTE GROUP

- ▶ Sparse Index
- ▶ Inverted Index
- ▶ None

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

- ▶ OLTP
- ▶ OLAP
- ▶ DSS

▶ inverted Index (Page 232)

351. Which of the following is not an "Orr's Law of Data Quality"?

- ▶ Data that is not used cannot be correct!"
- ▶ 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)

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

- ▶ "Define"
- ▶ Measure"
- ▶ Analyze"

▶ "Compression" (Page 194)

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

- ▶ OLTP
- ▶ OLAP
- ▶ Decision Support Systems
- ▶ None of these

354. Data mining, the _____ of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses.

▶ Extraction

▶ Insertion

AL-JUNAID INSTITUTE GROUP

▶ Inclusion

▶ Enclosure

355.) 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

▶ Drive

▶ File

▶ Log

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

▶ Mathematical

▶ Computational

▶ Statistical (Page 251)

▶ None of these

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

▶ OLTP (Page 254)

▶ OLAP

▶ DSS

▶ DWH

358. Question # 9 of 10 (Total M a r k s: 1) If 'M' rows from table-A match the conditions in the query then table-B is accessed 'M' times. Suppose table-B has an index on the join column. If 'a' I/Os are required to read the data block for each scan plus 'b' I/Os for each data block then the total cost of accessing table-B is _____ logical I/Os 358. approximately. Select correct option:

▶ $(a + b)M$ (Right Answer)

▶ $(a - b)M$

AL-JUNAID INSTITUTE GROUP

- ▶ $(a + b + M)$
- ▶ $(a * b * M)$

359. A data warehouse implementation without an OLAP tool is always possible.

- ▶ True
- ▶ False (Page 69)

360.) _____ modeling technique is more appropriate for data warehouses.

- ▶ entity-relationship
- ▶ dimensional
- ▶ physical
- ▶ None of the given

361. The performance in a MOLAP cube comes from the $O(1)$ look-up time for the array data structure.

- ▶ True (Page 79)
- ▶ False

362. Multi-dimensional databases (MDDs) typically use _____ formats to store pre-summarized cube structures.

- ▶ SQL
- ▶ proprietary file (Page 79)
- ▶ Object oriented
- ▶ Non- proprietary file

363.) Slice and Dice is changing the view of the data.

- ▶ True
- ▶ False (Page 81)

364. Data warehousing and on-line analytical processing (OLAP) are _____ elements of decision support system.

- ▶ Unusual
- ▶ Essential (Page 69)

AL-JUNAID INSTITUTE GROUP

▶ Optional

▶ None of the given

365. Virtual cube is used to query two similar cubes by creating a third "virtual" cube by a join between two cubes.

▶ True

▶ False (Page 86)

366. Analytical processing uses _____, instead of record level access.

▶ multi-level aggregates (Page 74)

▶ Single-level aggregates

▶ Single-level hierarchy

▶ None of the Given

367. The divide&conquer cube partitioning approach helps alleviate the _____ limitations of MOLAP

▶ implementation.

▶ Flexibility

▶ Maintainability

▶ Security

▶ Scalability (Page 85)

368. In a traditional MIS system, there is an almost linear sequence of queries.

▶ True (Page 72)

▶ False

369. Data Warehouse provides the best support for analysis while OLAP carries out the _____ task.

370. ▶ Mandatory

371. ▶ Whole

372. ▶ Analysis (Page 69)

373. ▶ Prediction

AL-JUNAID INSTITUTE GROUP

374. DOLAP allows download of "cube" structures to a desktop platform with the need for shared relational or cube server.

▶ True

▶ False (Page 78)

375. The STAR schema used for data design is a _____ consisting of fact and dimension tables. Select correct option:

▶ Network model

▶ Relational model (Page 87)

▶ Hierarchical data model

▶ None of the given

