1. Which of the following are the kinds of data warehouse applications
   1. Information processing
   2. Analytical processing
   3. Data Mining
   4. **All of the above**
2. Knowledge discovery from hidden patterns is typically called as \_\_\_\_\_\_\_\_\_.
   1. Information processing
   2. Analytical processing
   3. **Data Mining**
   4. None of the above
3. \_\_\_\_\_\_\_\_integrates together OLAP, data mining and mining knowledge in multidimensional databases.
   1. Data Mining
   2. Data Warehousing
   3. **On Line Analytical Mining**
   4. KDD
4. Which of the following choice/s are given to base cuboid in data cube materialization?
   1. No Materialization
   2. Full Materialization
   3. Partial Materialization
   4. **All of the above**
5. What is typical difference between Multidimensional database and Relational Database?
   1. **Multidimensional database stores data in more than 2 dimensions, whereas a relational database is a two-dimensional table.**
   2. Relational database stores data in more than 2 dimensions, whereas a Multidimensional database is a two-dimensional table.
   3. They are different names given to same type of database.
   4. None of the above
6. Using Hybrid OLAP is more advantageous since
   1. It gives faster performance like Multidimensional OLAP
   2. It provides Drill through capabilities of ROLAP
   3. It is very fast in retrieving the data.
   4. **All of the above**
7. \_\_\_\_\_\_\_\_\_\_ is a summarization of general features of objects in a target class and produces what is called characteristic rules.
   1. Data Materialization
   2. **Data Generalization**
   3. Data Mining
   4. Data Warehousing
8. A \_\_\_\_\_\_\_ represents a container for data values, a place where data values may be stored and later retrieved.
   1. Data Set
   2. Data Attributes
   3. **Data object**
   4. Structured Data
9. Which of the following are typically categorized as types of the attributes
   1. Nominal
   2. Binary
   3. Ordinal
   4. **All of the above**
10. In which of the following. Both outcomes are equally important?
    1. **Symmetric Binary attributes**
    2. Asymmetric Binary attributes
    3. Both A and B
    4. None of the above types
11. The most frequent number\_\_\_that is, the number that occurs the highest number of times in the set of the numbners.
    1. Median
    2. **Mode**
    3. Mean
    4. Standard Deviation
12. Which of the following method can be used to find Dissimilarity of Numeric Data
    1. Euclidean
    2. Manhattan
    3. Minkowski distances.
    4. **Any one of the above mentioned methods**
13. \_\_\_\_\_\_\_\_\_\_ is also called ab Chebyshev Distance.
    1. Euclidean
    2. Manhattan
    3. **Minkowski distances.**
    4. There is no such method
14. \_\_\_\_\_\_\_is a measure of similarity that can be used to compare documents or, to give a ranking to documents with respect to a given vector of query words.
    1. Distance measures
    2. **Cosine similarity**
    3. Graphical representations
    4. Data Visualization
15. Which of the following is typically called as a measure of spread for a random variable?
    1. Mean
    2. **Variance**
    3. Standard deviation
    4. Empirical mean
16. Which of the following distance metric is used in K-means algorithm?
    1. Jaccard coefficient
    2. **Euclidean Distance**
    3. Cosine Similarity
    4. None of the above
17. What is true about Data Visualization?
    1. Data Visualization is used to communicate information clearly and efficiently to users by the usage of information graphics such as tables and charts.
    2. Data Visualization helps users in analyzing a large amount of data in a simpler way.
    3. Data Visualization makes complex data more accessible, understandable, and usable.
    4. **All of the above**
18. Which of the following methods can be used to visualize the data
    1. Graphs
    2. Charts
    3. Maps
    4. **All of the above**
19. Which is the first step in the knowledge discovery process?
20. **Data cleaning**
21. Data selection
22. Data integration
23. Data transformation
24. Which one of the following is the correct application of the data mining?
    1. Fraud detection
    2. Corporate Analysis & Risk management
    3. Management and market analysis
    4. **All of the above**
25. Which of the following does not belong to data mining?
    1. Data archaeology
    2. Knowledge extraction
    3. Data exploration
    4. **Data transformation**
26. How many fact table Star Schema Contains?
    1. **One**
    2. Two
    3. Three
    4. Four
27. Which of the following operation can be applied on data warehouse?
    1. **Read only**
    2. Write only
    3. Both
    4. None of these
28. In Star Schema what type of relationship exists between a dimension table and fact table?
    1. **One to Many**
    2. Many to Many
    3. One to One
    4. All of the above
29. The OLAP stands for\_\_\_\_\_
    1. Online Advanced Processing
    2. Online Analytical Performance
    3. Online Analog Parameter
    4. **Online Analytical Processing**
30. Incorrect or invalid data is known as
    1. Missing data
    2. Outlier
    3. Changing data
    4. **Noisy data**
31. What is KDD in data mining?
    1. **Knowledge Discovery Database**
    2. Knowledge Discovery Data
    3. Knowledge Data Definition
    4. Knowledge data house
32. What is true about Data Mining?
    1. Data Mining is defined as the procedure of extracting information from huge sets of data
    2. Data mining also involves other processes such as Data Cleaning, Data Integration, Data Transformation
    3. Data mining is the procedure of mining knowledge from data.
    4. **All of the above**
33. What is the use of data cleaning?
    1. To remove the noisy data
    2. correct the inconsistencies in data
    3. transformations to correct the wrong data
    4. **All of the above**
34. Which one of the following refers to the binary attribute?
    1. The natural environment of a certain species
    2. Systems that can be used without knowledge of internal operations
    3. **This takes only two values. In general, these values will be 0 and 1, and they can be coded as one bit**
    4. All of the above
35. Let us suppose that you have the following two 4- dimensional word vectors for two words w1 and w2 respectively:

w1=(0.2,0.1,0.3,0.4) and w(0.3,0,0.2,0.5)

What is the cosine similarity between w1 and w2?

1. **0.948**
2. 0.832
3. 0
4. 0.5
5. Which of the following is a good alternative to the star schema?
   1. snow flake schema
   2. star schema
   3. star snow flake schema
   4. **fact constellation**
6. Which of the following OLAP techniques offers best latency?
   1. HOLAP
   2. MOLAP
   3. **ROLAP**
   4. DOLAP
7. What is true about Data Warehouse?
   1. The actual discovery phase of a knowledge discovery process
   2. The stage of selecting the right data for a KDD process
   3. **A subject-oriented integrated time variant non-volatile collection of data in support of management**
   4. None of these
8. Which attribute is used for Dimensionality reduction?
   1. composite attributes
   2. derived attributes
   3. relevant attributes
   4. **irrelevant attributes**
9. Which of the following is NOT example of ordinal attributes?
   1. Ordered numbers
   2. **Zip codes**
   3. Movie ratings
   4. Military ranks
10. The star schema is composed of \_\_\_\_\_\_\_\_\_\_ fact table.  
    **A. one.**  
    B. two.  
    C. three.  
    D. four.
11. Expansion for DSS in DW is   
    A**. Decision Support system**   
    B. Decision Single System  
    C. Data Storable System.  
    D. Data Support System.
12. Data warehouse is

|  |  |
| --- | --- |
| [A.](javascript:void(0);) | The actual discovery phase of a knowledge discovery process |
| [B.](javascript:void(0);) | The stage of selecting the right data for a KDD process |
| [C.](javascript:void(0);) | **A subject-oriented integrated time variant non-volatile collection of data in support of management** |
| [D.](javascript:void(0);) | None of these |

1. Data warehouse contains\_\_\_\_\_\_\_\_\_\_\_\_\_data that is never found in the operational environment.  
   A. normalized.  
   B. informational.  
   C. denormalized.  
   **D. summary.**
2. The data is stored, retrieved and updated in ……………

**A. OLTP**

B. OLAP

C. SMTP

D. FTP

42 What do data warehouses support?  
A. OLTP  
**B OLAP**  
C. OLAP and OLTP  
D. Operational databases

43 The core of the multidimensional model is the ....................... , which consists of a large set of facts and a number of dimensions.

**A.** [**Data cube**](http://en.wikipedia.org/wiki/OLAP_cube)

B. Dimensions cube

C .Multidimensional cube

D. Data model

44.  .......................... supports basic OLAP operations, including slice and dice, drill-down, roll-up and pivoting.

A. Information processing

B. Transaction processing

C. Data mining

**D. Analytical processing**

46. Which of the following is not a data pre-processing methods

**A. Data Visualization**

B. Data Discretization

C. Data Cleaning

D. Data Reduction

47. Dimensionality reduction reduces the data set size by removing

A. composite attributes

B. derived attributes

C. relevant attributes

**D. irrelevant attributes**

48. Incorrect or invalid data is known as

a. Missing data

b. Outlier

c. Changing data

**d. Noisy data**

49. Synonym for data mining is

A. Data Warehouse

B. OLAP

C. Business intelligence

**D. Knowledge discovery in database**

50. Binary attribute are

|  |  |
| --- | --- |
| [A.](javascript:void(0);) | Contains order |
| [B.](javascript:void(0);) | **This takes only two values. In general, these values will be 0 and 1 and .they can be coded as one bit** |
| [C.](javascript:void(0);) | Systems that can be used without knowledge of internal operations |
| [D.](javascript:void(0);) | None of these |

51 What is the out put of KDD…

A. Query

B. Data

**C. Useful Information**

D. Chart

52. Which of the following is not belong to data mining?

**A . Data transformation**

B. Knowledge extraction

C. Data exploration

D. Data archaeology

53**.** [Data mining](https://t4tutorials.com/what-is-data-mining-what-is-not-data-mining/)is?

A. time variant non-volatile collection of data

B. The stage of [selecting the right data](https://t4tutorials.com/which-attribute-selection-measure-is-best-in-data-mining/)

C.The actual discovery phase of a knowledge

D. None of these