

**U.G. Department of Computer Applications**

**N.G.M College**

**16 UBC 626 - Data Mining And Warehousing**

**Multiple Choice Questions. (K1 Questions)**

**UNIT - I**

1. \_\_\_\_\_ is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management decisions.

- A. Data Mining.                      B. Data Warehousing.                      C. Web Mining                      D. Text Mining

2. The data Warehouse is \_\_\_\_\_.

- A. Read only.                      B. Write only.                      C. Read write only.                      D. None.

3. The important aspect of the data warehouse environment is that data found within the Data Warehouse is \_\_\_\_\_.

- A. Subject-oriented.                      B. Time-variant.  
C. Integrated.                      D. All of the above.

4. The time horizon in Data warehouse is usually \_\_\_\_\_.

- A. 1-2 years.    B. 3-4 years.    C. 5-6 years.    D. 5-10 years.

5. The data is stored, retrieved & updated in \_\_\_\_\_.

- A. OLAP.                      B. OLTP.                      C. SMIP.                      D. FIP.

6. \_\_\_\_\_ describes the data contained in the data warehouse.

- A. Relational data.                      B. Operational data.  
C. Metadata.                      D. Informational data.



15. Bill Inmon has estimated \_\_\_\_\_ of the time required to build a data warehouse, is consumed in the conversion process.

- A. 10 percent.      B. 20 percent.      C. 40 percent      D. 80 percent.

16. Detail data in single fact table is otherwise known as \_\_\_\_\_.

- A. Monoatomic data.      B. Diatomic data.      C. Atomic data.      D. Multiatomic data.

17. \_\_\_\_\_ test is used in an online transactional processing environment.

- A. MEGA.      B. MICRO.      C. MACRO.      D. ACID.

18. \_\_\_\_\_ is a good alternative to the star schema.

- A. Star schema. B. Snowflake schema. C. Fact constellation. D. Star-snowflake schema.

19. The biggest drawback of the level indicator in the classic star-schema is that it limits \_\_\_\_\_.

- A. Quantify.      B. Qualify.      C. Flexibility.      D. Ability.

20. A data warehouse is \_\_\_\_\_.

- A. Updated by end users.      B. Contains numerous naming conventions and formats  
C. Organized around important subject areas.      D. Contains only current data.

### UNIT - III

21. Query tool is meant for \_\_\_\_\_.

- A. Data acquisition.      B. Information delivery.      C. Information exchange.      D. Communication.

22. The full form of KDD is \_\_\_\_\_.

- A. Knowledge Database.      B. Knowledge Discovery in Database.  
C. Knowledge Data house.      D. Knowledge Data Definition.

23. The first International conference on KDD was held in the year \_\_\_\_\_.

- A. 1996.      B. 1997.      C. 1995      D. 1994.

24. Removing duplicate records is a process called \_\_\_\_\_.

- A. Recovery.                      B. Data cleaning.                      C. Data cleansing.                      D. Data pruning.

25. \_\_\_\_\_ contains information that gives users an easy-to-understand perspective of the information stored in the data warehouse.

- A. Business metadata.    B. Technical metadata.    C. Operational metadata.    D. Financial metadata.

26. \_\_\_\_\_ helps to integrate, maintain and view the contents of the data warehousing system.

- A. Business directory.                      B. Information directory.  
C. Data dictionary.                      D. Database.

27. Discovery of cross-sales opportunities is called \_\_\_\_\_.

- A. Segmentation.                      B. Visualization.  
C. Correction.                      D. Association.

28. Data marts that incorporate data mining tools to extract sets of data are called \_\_\_\_\_.

- A. Independent data mart.                      B. Dependent data marts.  
C. Intra-entry data mart.                      D. Inter-entry data mart.

29. \_\_\_\_\_ can generate programs itself, enabling it to carry out new tasks.

- A. Automated system.                      B. Decision making system.  
C. Self learning system.                      D. Productivity system.

30. The power of self-learning system lies in \_\_\_\_\_.

- A. Cost.                      B. Speed.                      C. Accuracy.                      D. Simplicity.

#### UNIT - IV

31. Building the informational database is done with the help of \_\_\_\_\_.

- A. Transformation or propagation tools. B. Transformation tools only.  
C. Propagation tools only. D. Extraction tools.

32. How many components are there in a data warehouse?

- A. Two. B. Three. C. Four. D. Five.

33. Which of the following is not a component of a data warehouse?

- A. Metadata. B. Current detail data.  
C. Lightly summarized data. D. Component Key.

34. \_\_\_\_\_ is data that is distilled from the low level of detail found at the current detailed level.

- A. Highly summarized data. B. Lightly summarized data.  
C. Metadata. D. Older detail data

35. Highly summarized data is \_\_\_\_\_.

- A. Compact and easily accessible. B. Compact and expensive.  
C. Compact and hardly accessible. D. compact.

36. Metadata contains atleast \_\_\_\_\_.

- A. The structure of the data. B. The algorithms used for summarization.  
C. The mapping from the operational environment to the data warehouse.  
D. All of the above.

37. \_\_\_\_\_ can generate programs itself, enabling it to carry out new tasks.

- A. Automated system. B. Decision making system. C. Self-learning system. D. Productivity system.

38. The power of self-learning system lies in \_\_\_\_\_.

- A. Cost. B. Speed. C. Accuracy. D. Simplicity.

39. How many components are there in a data warehouse?

A. Two.                      B. Three.                      C. Four.                      D. Five.

40. \_\_\_\_\_ employs the supervised mode of learning.

A. RBF.                      B. MLP.                      C. MLP & RBF.                      D. ANN.

## UNIT - V

41. The actual amount of reduction at each learning step may be guided by \_\_\_\_\_.

A. Learning cost.                      B. LearninLevel.                      C. Learning Rate                      . D. Learning Time.

42. The terms equality and roll up are associated with \_\_\_\_\_.

A. OLAP.                      B. Visualization.                      C. Data mart.                      D. Decision tree

43. Treating incorrect or missing data is called as \_\_\_\_\_.

A. Selection.                      B. Preprocessing.                      C. Transformation.                      D. Interpretation.

44. Converting data from different sources into a common format for processing is called as \_\_\_\_\_.

A. Selection.                      B. Preprocessing.                      C. Transformation.                      D. Interpretation.

45. Various visualization techniques are used in \_\_\_\_\_ step of KDD.

A. Selection.                      B. Transformation.                      C. Data mining.                      D. Interpretation.

46. Exceptional reporting in data warehousing is otherwise called as \_\_\_\_\_.

A. Exception.                      B. Alerts.                      C. Errors.                      D. Bugs.

47. \_\_\_\_\_ is used to proceed from very specific knowledge to more general information.

A. Induction.                      B. Compression.                      C. Approximation.                      D. Substitution.

48. Describing some characteristics of a set of data by a general model is viewed as \_\_\_\_\_.

A. Induction.                      B. Compression.                      C. Approximation.                      D. Summarization.

49. \_\_\_\_\_ helps to uncover hidden information about the data.

- A. Induction.            B. Compression.            C. Approximation.            D. Summarization.

50. Incorrect or invalid data is known as \_\_\_\_\_.

- A. Changing data.            B. Noisy data.            C. Outliers.            D. Missing data.

### Answer Key

1. B    2. A    3. D    4. D    5. B    6. C    7. B    8. B    9. D    10. B  
11. D    12. A    13. C    14. A    15. D    16. C    17. D    18. C    19. C    20. C  
21. A    22. B    23. C    24. B    25. A    26. B    27. D    28. B    29. D    30. C  
31. A    32. D    33. D    34. B    35. A    36. D    37. D    38. C    39. D    40. C  
41. C    42. C    43. B    44. C    45. D    46. B    47. A    48. B    49. C    0. B

**17 UBC 626 DATA MINING AND WAREHOUSING  
(5 Marks Questions)**

**Unit I**

1. What is Data Mining? Explain.
2. Explain Information as a Production Factor.
3. Give details about Data Mining in Marketing.
4. Define Learning.
5. Explain Self-Learning Computer Systems.
6. Explain types of Machine Learning.
7. Write a note on Concept Learning?

**Unit II**

1. Compare Data Mining and Data Warehousing.
2. What is the Need of Data Warehouse?
3. Explain Integration with Data Mining.
4. Explain the types of Multiprocessing Machines.
5. Write a short note on Expert system?

**Unit III**

1. What is Data Selection?
2. What is Cleaning?
3. What is Enrichment?
4. What is Coding?
5. What is Data Mining?
6. Explain Data Visualization Techniques.
7. Explain in detail about Likelyhood and Distance.
8. Explain in detail about Association Rules.
9. Explain in detail about Neural Networks.
10. Explain in detail about Genetic Algorithms.
11. Explain in detail about Reporting.

**Unit IV**

1. Explain the different forms of Knowledge in detail.
2. What do you mean by Data Selection? Explain.
3. What do you mean by Cleaning? Explain.
4. What do you mean by Enrichment? Explain.
5. What do you mean by Coding? Explain.
6. Explain in detail about Reporting.



## **Unit V**

1. Discuss about the Contents of a Message.
2. Explain Noise and Redundancy.
3. Discuss about the Significance of Noise.
4. Explain the concept - From Keys of Statistical Dependencies.
5. Explain about the Demoralization.

## **17 UBC 626 DATA MINING AND WAREHOUSING (8 Marks Questions)**

### **Unit I**

1. Differentiate Data Mining Vs Query Tools.
2. Explain briefly about practical Applications of Data Mining.
3. Define briefly about the concepts of Learning.
4. Explain Machine Learning and Methodologies of Science.
5. Discuss the Issues of Concept Learning.

### **Unit II**

1. Define Data Warehouse and their types in detail.
2. Difference between OLAP and OLTP.
3. Explain detail about Meta data.
4. Explain briefly about Designing a Decision Support Systems.
5. Discuss about Client Server and Data Warehousing.
6. Give detail explanation about Multiprocessing Machines.
7. Define Cost Justification.

### **Unit III**

1. Explain briefly about Knowledge Discovery Process.
2. Give Preliminary Analysis of Data Set Using Relational Query Tools.
3. Explain Data Visualization Techniques.
4. Explain in detail about OLAP Tools.
5. Explain in detail about K-Nearest Neighborhood.
6. Explain in detail about Decision Trees.

### **Unit IV**

1. Setting Up KDD Environment: Explain.
2. Explain in detail about Data Selection and Cleaning
3. Explain in detail about Enrichment and Coding
4. Explain in detail about role of Reporting in Data warehouse.
5. What are the 10 Golden Rules? Explain.

## **Unit V**

1. Discuss Some Formal aspects of Learning.
2. Explain about the learning of Comprehension of Data Sets.
3. Explain about Noise and Redundancy and give the Significance of Noise.
4. Explain about the Fuzzy Database.
5. Traditional Theory of Relational Database from Relations of Tables.
6. What are Data Mining Primitives? Explain.

