

- b. Explain frequent pattern mining without candidate generation. 12 2 4 1
32. a. Explain in detail about HARBOR – A highly available data warehouse. 12 2 5 1
- (OR)
- b. Explain with a illustration about data warehouse in Hewlett Packard. 12 2 5 1

\*\*\*\*\*

Reg. No.															
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

**B.Tech. DEGREE EXAMINATION, MAY 2023**  
Fifth & Sixth Semester

**18CSE487T – DATA WAREHOUSING AND ITS APPLICATION**  
(For the candidates admitted during the academic year 2018-2019 to 2021-2022)

**Note:**

- (i) **Part - A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40<sup>th</sup> minute.
- (ii) **Part - B & Part - C** should be answered in answer booklet.

Time: 3 hours

Max. Marks: 100

**PART – A (20 × 1 = 20 Marks)**

Answer **ALL** Questions

- |  | Marks | BL | CO | PO |
|--|-------|----|----|----|
| 1. Which one of the following is not a type of data warehousing architecture?<br>(A) Centralized (B) Distributed<br>(C) Hybrid (D) Unstructured  | 1     | 1  | 1  | 1  |
| 2. What are the benefits of using a data mart?<br>(A) Increased data security (B) Faster access to relevant data<br>(C) Improved data quality (D) Improved data accuracy   | 1     | 2  | 1  | 1  |
| 3. DBA functions are handled by _____.<br>(A) DW administrator (B) Lead architect<br>(C) Data modeler (D) Business analyst   | 1     | 2  | 1  | 1  |
| 4. The requirement definition document should include the following information except _____.<br>(A) Available data source (B) Location of the data source<br>(C) Data extraction procedures (D) No data structure information                                   | 1     | 1  | 1  | 1  |
| 5. Data staging will perform _____.<br>(A) Data cleaning (B) DBMS feature<br>(C) Growth potential (D) External data source   | 1     | 2  | 1  | 1  |
| 6. Which of the following is an advantage of using a dimensional model for data warehousing?<br>(A) It is easier to maintain and update (B) It requires less storage space<br>(C) It is optimized for transaction processing (D) It provides a normalized schema | 1     | 2  | 2  | 1  |
| 7. Which of the following is a characteristic of a star schema in a dimensional model?<br>(A) It contains multiple hierarchies (B) It is normalized<br>(C) It has a single, central fact table (D) It contains many-to-many relationship                         | 1     | 2  | 2  | 1  |

8. Which of the following is not a meta data management tool? 1 2 3 1  
(A) Data dictionaries (B) Ontologies  
(C) Data warehouse (D) Taxonomies
9. What is OLAP used for? 1 2 2 1  
(A) Data entry (B) Data storage  
(C) Data analysis (D) Data processing
10. What is dimension in dimensional analysis? 1 2 3 1  
(A) A measurement unit (B) A categorical variable used for data analysis  
(C) A numerical variable used for data analysis (D) A data visualization tool
11. What is the purpose of metadata query generation? 1 2 3 1  
(A) To extract information from metadata (B) To create metadata from raw data  
(C) To standardize metadata (D) To store metadata
12. What is recovery? 1 2 3 1  
(A) The process of creating a copy of data for protection (B) The process of moving data to another location  
(C) The process of restoring data after a failure (D) The process of deleting data
13. What is a decision tree in data mining? 1 2 4 1  
(A) A data visualization tool (B) A clustering technique  
(C) A regression analysis technique (D) A predictive modeling technique
14. Which of the following is a technique used in data transformation? 1 1 4 1  
(A) Scaling (B) Outlier removal  
(C) Data integration (D) Data cleaning
15. The left hand side of an association rule is called \_\_\_\_\_. 1 2 4 1  
(A) Consequent (B) On set  
(C) Antecedent (D) Precedent
16. Which of the following is required by K-means clustering? 1 1 4 1  
(A) Defined distance metric (B) Number of clusters  
(C) Initial guess as to cluster centroids (D) Assignment of each point to clusters
17. The \_\_\_\_\_ data management platform is a powerful platform for data transfer, transformation and automation 1 1 5 1  
(A) Network (B) Y2B  
(C) Clover Dx (D) Rank Dx
18. The business driven data warehouse automation solution is \_\_\_\_\_. 1 1 5 1  
(A) Data vault creator (B) Informatica

- (C) Panoply (D) TiltMx
19. EDEKA will be able to instantly share product information throughout their supply chain and more than 6,000 stores by 1 2 5 1  
(A) Hybrid architecture (B) Action driven architecture  
(C) Event driven architecture (D) Network architecture
20. The simplest approach to sync, store and retrieve your data is using \_\_\_\_\_, 1 1 5 1  
which combines a data warehousing with built-in data connections.  
(A) Panoply (B) Monopoly  
(C) Storage (D) Access

**PART – B (5 × 4 = 20 Marks)**  
Answer ANY FIVE Questions

Marks BL CO PO

21. A data warehouse in an environment, not a product. Discuss. 4 1 1 1
22. What are the various sources for the data warehouse? Explain. 4 2 1 1
23. Explain the sparse nature of the data in the fact table. 4 2 2 1
24. Differentiate between slowly and rapidly changing dimensions. 4 2 3 1
25. Name five types of major transformation tasks. 4 1 4 1
26. Explain data integration in brief. 4 1 4 1
27. Explain confidence metrics to evaluate the strength of association rule. 4 1 4 1

**PART – C (5 × 12 = 60 Marks)**  
Answer ALL Questions

Marks BL CO PO

28. a. Discuss about data warehouse architecture components with neat diagram. 12 1 1 1
- (OR)
- b. Explain data warehouse development phase with neat sketch. 12 1 1 1
29. a. Discuss any two data extraction technique with a suitable sketch. 12 1 2 1
- (OR)
- b. Explain dimension analysis with proper example. 12 1 2 1
30. a. Explain the building and implementing of data marts. 12 1 3 1
- (OR)
- b. Describe the working of backup and recovery of data warehousing. 12 1 3 1
31. a. Explain Apriori algorithm with suitable example. 12 1 4 1

(OR)