b.	Explain snowflake schema technique with an example. Mention its advantages and disadvantages.	10	2	2	3
28. a.i.	Identify critical success factors of data warehouse environment.	6	2	2	3
ii.	Give the guidelines for building a successful warehouse.	4	1	2	3
	(OR)				
b.	Explain about meta data management and query management in detail.	10	1	2	2
29. a.i.	Write short note on functionality of data mining.	5	2	5	4
ii.	Describe about association rule mining with an example.	5	1	5	4
	(OR)				
b.	Write briefly about data mining application.	10	2	5	5
30. a.	Give the case study for EDEKA data warehouse.	10	3	6	5
	(OR)				
b.	Brief about the HARBOR data warehouse.	10	3	6	5

* * * * *

Reg. No.														

B.Tech. DEGREE EXAMINATION, NOVEMBER 2022

Sixth/ Seventh Semester

18CSE487T – DATA WAREHOUSING AND ITS APPLICATIONS

(For the candidates admitted from the academic year 2018-2019 to 2019-2020)

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 40th minute.

(ii) **Part - B** should be answered in answer booklet.

Time: 2½ Hours Max. Marks: 75

		Marks	BL	CO	PO
	PART – A (25 × 1 = 25 Marks) Answer **ALL** Questions				
1.	_____ is the process of removing errors from input data. (A) Data extraction (B) Data cleansing (C) Data warehousing (D) Data mining	1	1	2	3
2.	_____ layer is involved in scheduling various task that must be accomplished to maintain data and metadata in the data warehouse. (A) Process management (B) Application messaging (C) Data staging (D) Data access	1	1	2	3
3.	In _____ approach, data is extracted from the operational systems, transformed, cleaned and integrated to finally store in data warehouse. (A) Centralized (B) Distributed (C) Top down (D) Bottom up	1	1	2	3
4.	_____ architecture is highly scalable and provides fast access between nodes. (A) Massively parallel processing (B) Cluster (C) Symmetric multi processing (D) Cache coherent non uniform memory	1	1	2	3
5.	In data quality tool _____ feature improve merging of data from dissimilar data sources. (A) Error checking (B) Error correction (C) Error detection (D) Error discovery	1	1	2	3
6.	_____ is a design technique to structure the business dimension and analyze the metrics. (A) Operational database modelling (B) Dimensional modelling (C) ER model (D) Data warehouse modelling	1	1	4	3
7.	How many fact tables are there in star schema? (A) 1 (B) 2 (C) 4 (D) 3	1	1	4	3

8. _____ performs aggregation of data by reduction in dimension (or) by stepping up a concept hierarchy for a dimension. 1 1 1 3
(A) Drill down (B) Slice
(C) Roll up (D) Dice

9. _____ is an area where the data is hold temporary on data warehouse server. 1 1 4 3
(A) Data purging (B) Data warehouse
(C) Data capture (D) Data staging

10. The commercial product Microsoft analysis services is an example of 1 1 1 3
(A) Desktop OLAP model (B) Relational OLAP model
(C) Multidimensional OLAP model (D) Hybrid OLAP model

11. The _____ operation has to be done at the end of the backup process. 1 1 4 3
(A) Yearly (B) Last overnight
(C) Day to day (D) Hourly

12. To implement security features, you may use security mechanism at how many levels? 1 1 4 3
(A) 4 (B) 3
(C) 1 (D) 2

13. _____ can be divided into production reporting tools and desktop report writers. 1 1 2 3
(A) Access tools (B) OLAP tools
(C) Reporting tools (D) Query tools

14. _____ is important because you can leverage meta data in understanding, aggregating, grouping and sorting data for use 1 1 4 3
(A) User management (B) Query management
(C) Data management (D) Metadata management

15. In _____ operation, older data is rolled up into aggregated form to reduce the space needed to store the entire data. 1 1 1 3
(A) Data rollup (B) Data cleansing
(C) Data gathering (D) Data transformation

16. _____ refers to the technique of searching useful and relevant information from the data warehouse. 1 1 4 3
(A) Data mart (B) Data modelling
(C) Data gathering (D) Data mining

17. _____ helps to label the variables without providing the numerical value. 1 1 4 3
(A) Discrete data (B) Continuous data
(C) Ordinal data (D) Nominal data

18. In the pursuit of seeking knowledge from data, _____ process is highly iterative and interactive. 1 1 5 2
(A) Data warehousing (B) Knowledge discovery
(C) Data mining (D) Data mart

19. Which storage of knowledge discovery involves the combination of multiple data sources? 1 1 4 3
(A) Data selection (B) Data transformation
(C) Data cleaning (D) Data integration

20. _____ is a non parametric test that measures the strength of dependence between two variables 1 1 5 3
(A) Kendall rank correlation (B) Spearman rank correlation
(C) Pearson correlation (D) Positive correlation

21. _____ is the process of transforming unstructured text into structural format to identify meaningful patterns 1 1 6 5
(A) Spatial data mining (B) Web mining
(C) Text mining (D) Data mining

22. Data warehouse technology provided to the department of _____ by NIC eliminate high response time problems by storing current and historical data from disparate information systems. 1 1 6 5
(A) Finance (B) Treasuries
(C) Ministry (D) Education

23. Hewlett Packard has at least _____ OLAP cubes, each of will support a particular group of decision makers 1 1 6 5
(A) 4 (B) 8
(C) 6 (D) 2

24. The retail industry in Germany _____ felt that it provides the ability to conduct analyses on information such as sales turnover and inventor levels. 1 1 6 5
(A) IKEA (B) ZARA
(C) ALDI (D) EDEKA

25. The first version of _____ of Egg was built on sun fire 6800 server. 1 1 6 5
(A) Customer data warehouse (B) Real application cluster
(C) Storage area network (D) Data base administrator

PART – B (5 × 10 = 50 Marks) Marks BL CO PO
Answer **ALL** Questions

26. a. Explain about the common architecture of data warehouse. 10 1 1 3

(OR)

b.i. Summarize about data warehouse readiness assessment. 6 3 4 4

ii. How is the project team organized to build a data warehouse project? 4 2 4 4

27. a.i. Explain dimensional modeling. 5 1 1 2

ii. Write short notes on the type of transformation task performed on extracted data. 5 2 1 2

(OR)