Reg. No. : | E | N | G | G | T | R | E | E | . | C | O | M |

Question Paper Code : 40442

B.E./B.Tech. DEGREE EXAMINATIONS, NOVEMBER/DECEMBER 2024.

Fifth/Sixth Semester

Computer Science and Engineering

CCS 341 – DATA WAREHOUSING

(Common to : Computer Science and Design/ Computer Science and Engineering (Artificial Intelligence and Machine Learning)/ Computer Science and Engineering (Cyber Security)/ Computer and Communication Engineering/ Artificial Intelligence and Data Science/ Computer Science and Business Systems/ Information Technology)

(Regulations 2021)

Time : Three hours Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. Write the different steps in knowledge discovery in databases.

2. How is data warehouse different from a database? Identify the similarity.

3. What is ETL?

4. Outline the characteristics of OLAP.

5. Compare DataMart with Data Warehouse.

6. Why is a data mart considered cost-effective compared to a data warehouse?

7. Name the types of data warehouse schema.

8. What is the significance of a fact constellation schema in dimensional modelling?

9. Recall the responsibilities of a Data Warehousing System Configuration Manager.

10. Outline the primary task of a Data Warehousing System Backup Recovery Manager.

PART B — (5 × 13 = 65 marks)

11. (a) (i) With a neat sketch, explain the Steps for design and construction of Data warehouses and explain with three tier architecture. (6)

(ii) Describe the three layers of Snowflake's architecture. (7)

Or

(b) (i) Suppose that a data warehouse consists of four dimensions customer, product, salesperson, and sales time and the three measure sales Amount (in rupees), VAT (in rupees) and payment type (in rupees). Draw the different classes of schemas that are popularly used for modeling data warehouses and explain it. (8)

(ii) Discuss on autonomous data warehouse. (5)

12. (a) (i) Compare OLAP with OLTP system. (6)

(ii) Discuss the typical OLAP operations with an example (KDD). (7)

Or

(b) (i) Diagrammatically illustrate and describe the architecture of MOLAP, ROLAP, HOLAP. (7)

(ii) Identify the Major differences between MOLAP and ROLAP. (6)

13. (a) Explain the following

(i) Metadata repository. (7)

(ii) Role of Metadata. (6)

Or

(b) (i) Compare and contrast the advantages and disadvantages of vertical partitioning and horizontal partitioning in a data warehousing context. (7)

(ii) Describe the challenges of metadata management. (6)

14. (a) (i) Propose a fact constellation schema for a healthcare data warehouse to support complex analytical queries. (7)

(ii) Describe the process architecture involved in designing and implementing a star schema. (6)

Or

(b) (i) Explain how does a snowflake schema differ from a star schema in terms of normalization? (7)

(ii) Explain the relationship between the Load Manager and the Warehouse Manager in a data warehousing process. (6)

15. (a) Explain about various data warehousing process managers, its collaboration and interaction to ensure operations and performance of the entire data warehousing system. (13)

Or

(b) Justify the statement "Testing process contribute to the overall quality a data warehousing system", and describe the types of tests conducted on the system. (13)

PART C — (1 × 15 = 15 marks)

16. (a) (i) Design a multidimensional cube with an example. (5)

(ii) Suppose that a data warehouse consists of the four dimensions date, spectator, location and game, and the two measures count and charge, where charge is the fare that a spectator pays when watching a game on a given date spectators may be students, adults, or seniors, with each category having its own charge rate.

(1) Draw a star schema diagram for the data warehouse. (4)

(2) Starting with a base cuboid [date, spectator, location, game], what specific OLAP operations should one perform in order to list the total charge paid by student spectators at GM_Place in 2000? (6)

Or

(b) Consider you are working for a retail chain that operates in multiple regions and sells a variety of products both in physical stores and online. The company wants to improve its decision-making process by analyzing sales data more effectively. They have specific business requirements to be met through the design of a data mart.

Requirements:

(i) Analyze sales performance across different regions, product categories and sales channels (physical stores vs. online).

(ii) Identify trends in customer purchasing behavior, including popular products, seasonal trends and customer demographics.

(iii) Track inventory levels and monitor stock movement to optimize inventory management and prevent stockouts.

(iv) Integrate data from various sources, including sales transactions, inventory databases and customer demographics, to provide a comprehensive view of the business.

Design a data mart to meet the above mentioned requirements.
