

NOTE:

1. Answer question 1 and any FOUR from questions 2 to 7.
2. Parts of the same question should be answered together and in the same sequence.

Time: 3 Hours**Total Marks: 100**

1.
 - a) What is ETL process? Explain each step.
 - b) Why machine learning is done?
 - c) What is the difference between “supervised and unsupervised” learning scheme.
 - d) Explain some of the data mining applications?
 - e) What are data cubes? Explain how to mine an OLAP cube.
 - f) What is an outlier? Mention its application.
 - g) What is Data Mart?

(7x4)

2.
 - a) Explain the architecture of data warehouse and explain its components with diagram.
 - b) Define Data warehouse. Discuss the characteristics or features of data warehouse with example.

(9+9)

3.
 - a) What is dimensional modelling? Discuss Star schema in detail. Also differentiate between star schema and snowflake schema. State three advantages and disadvantages.
 - b) Explain the various OLAP operations in detail

(12+6)

4.
 - a) Create star schema corresponding to order analysis. Explain its dimension table and fact table. Also discuss the dimension hierarchies.
 - b) Discuss the characteristics of fact table and dimension table.
 - c) Explain type 1, type 2, type 3 changes in data warehouse while performing dimensional modelling.

(6+3+9)

5.
 - a) Explain OLAP and its models? Also differentiate between ROLAP, MOLAP and HOLAP.
 - b) Distinguish between Online Transaction Processing (OLTP) and Online Analytical Processing.

(12+6)

6.
 - a) Define Knowledge discovery. What are Knowledge discovery phases? Discuss in detail with diagram.
 - b)
 - i) What is Descriptive and predictive data mining?
 - ii) What are the requirements of clustering?

(9+6+3)

7. Write short notes on **any three** of the followings:
 - a) Classification
 - b) Association Rule Mining
 - c) Decision Tress
 - d) Link Analysis

(6+6+6)