

N.B. : (1) Question No. 1 is **compulsory**.

(2) Solve any **four** out of the **remaining**.

(3) Draw suitable diagrams wherever **necessary**.

(4) Assume **suitable** data (if required)



1. (a) Define a data warehouse. Explain what is the need for developing a data warehouse and hence explain its architecture. **10**
 (b) Compare OLTP and OLAP systems. Explain the steps in KDD with a suitable block diagram. **10**
2. (a) What is meant by ETL ? Explain the ETL process in detail. **10**
 (b) State and explain the various schemas used in data warehousing with examples for each of them. **10**
3. (a) Differentiate between top down and bottom-up approaches for building a data warehouse. Explain the advantages and disadvantages of each of them. **10**
 (b) Define what is meant by information package diagram. For recording the information requirements for "hotel occupancy" having dimensions like time, hotel etc, give the information package diagram for the same, also draw the star schema and snow flake schema. **10**
4. (a) What is meant by meta data ? Explain with an example. Explain the different types of meta data stored in a data warehouse. **10**
 (b) Explain what is meant by association rule mining. For the table given below perform apriori algorithm. Also – **10**
 - (i) Determine the k-item sets (frequent) obtained.
 - (ii) Justify the strong association rule that has been determined i.e. specify which is the strongest rule obtained.

The table is as follows -

TID	Items
01	1, 3, 4, 6
02	2, 3, 5, 7
03	1, 2, 3, 5, 8
04	2, 5, 9, 10
05	1, 4

Assume Minimum support of 30% and
Minimum confidence of 75%.

5. (a) Explain dimension modelling in detail. 10
(b) Explain what is meant by clustering. State and explain the various types with suitable example for each. 10
6. (a) What is meant by classification ? Justify why clustering is said to be supervised learning. How is the classifier accuracy determined and also explain its various types. 10
(b) What is meant by market-basket analysis ? Explain with an example. State and explain with formula the meaning of the terms :- 10
(i) Support
(ii) Confidence
(iii) Iceberg queries.
Hence explain how to mine multi level association rules from transaction databases, with example for each.
7. Write short notes on (any two) :- 20
(a) OLAP operations
(b) Data warehouse deployment and maintenance
(c) Attribute oriented induction
(d) Web mining.