Time: 03 Hours **Marks: 80**

Note: 1. Question 1 is compulsory
2. Answer any three out of remaining five questions.
3. Assume any suitable data wherever required and justify the same.

Q1 a) What is dimensional modeling? Design the data warehouse for wholesale furniture Company. The data warehouse has to allow to analyze the company's situation at least with respect to the Furniture, Customer and Time. Moreover, the company needs to analyze: The furniture with respect to its type, category and material. The customer with respect to their spatial location, by considering at least cities, regions and states. The company is interested in learning the quantity, income and discount of its sales. **[10]**

b) i. Explain the architecture of data mining. **[10]**
ii. Explain different steps involved in data processing.

Q2 a) Differentiate top-down and bottom-up approaches for building data warehouse. Discuss the merits and limitations of each approach. **[10]**

b) Explain frequent pattern growth mining with example. **[10]**

Q3 a) For the following dataset, apply decision tree classification algorithm and show the generated rules **[10]**

Id	Age	Income	Student	Credit-rating	buys computer
1	young	high	no	fair	no
2	young	high	no	good	no
3	middle	high	no	fair	yes
4	old	medium	no	fair	yes
5	old	low	yes	fair	yes
6	old	low	yes	good	no
7	middle	low	yes	good	yes
8	young	medium	no	fair	no
9	young	low	yes	fair	yes
10	old	medium	yes	fair	yes
11	young	medium	yes	good	yes
12	middle	medium	no	good	yes
13	middle	high	yes	fair	yes
14	old	medium	no	good	No

b) Explain steps of ETL process in detail **[10]**

Q4 a) What is Clustering Techniques? Discuss the Agglomerative algorithm using following data and plot a Dendrogram using single link and complete link approach. The following figure contains sample data items indicting the distance between the elements: **[10]**

Item	E	A	C	B	D
E	0	1	2	2	3
A	1	0	2	5	3
C	2	2	0	1	6
B	2	5	1	0	3
D	3	3	6	3	0

b) i. Explain different OLAP models. **[10]**

ii. Differentiate Online transaction processing (OLTP) and Online analytical processing (OLAP)

Q5 a) Consider a data warehouse for a hospital where there are three dimension a) Doctor b) Patient c) Time. Consider a measure charge fee that the doctor charges to a patient for a visit. Create a cube and illustrate the following OLAP operations: **[10]**

1) Rollup 2) Drill down 3) Slice 4) Dice 5) Pivot.

2)

b) Discuss Association Rule Mining (AR) and Apriori Algorithm. Apply AR Mining to find all frequent item sets and association rules for the following dataset: **[10]**

Q6 Write short notes on any *two* of the following: **[20]**

a) Linear Regression

b) Data Visualization

c) DBSCAN clustering
