Q1 a) Information requirements are recorded for “Hotel occupancy” considering [10] dimensions like Hotel, Room and Time. Few Facts recorded are vacant rooms, occupied rooms, number of occupants, etc. Answer the following questions for this problem:
   i. Design the star schema.
   ii. Can you convert this star schema to a snowflake schema? If yes, justify and draw the snowflake schema.

b) Explain Data mining as a step in KDD. Illustrate the architecture of typical data mining system.

Q2 a) The college wants to record the Marks for the courses completed by students using [10] the dimensions: I) Course, II) Student, III) Time & a measure Aggregate marks. Create a Cube and perform following OLAP operations:
   i) Rollup
   ii) Drill down
   iii) Slice
   iv) Dice
   v) Pivot.

b) Apply the Naive Bayes classifier algorithm to classify an unknown sample X (outlook = sunny, temperature = cool, humidity = high, windy = false)
The sample data set is as follows:

<table>
<thead>
<tr>
<th>Outlook</th>
<th>Temperature</th>
<th>Humidity</th>
<th>Windy</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunny</td>
<td>Hot</td>
<td>High</td>
<td>False</td>
<td>N</td>
</tr>
<tr>
<td>Sunny</td>
<td>Hot</td>
<td>High</td>
<td>True</td>
<td>N</td>
</tr>
<tr>
<td>Overcast</td>
<td>Hot</td>
<td>High</td>
<td>False</td>
<td>P</td>
</tr>
<tr>
<td>Rain</td>
<td>Mild</td>
<td>High</td>
<td>False</td>
<td>P</td>
</tr>
<tr>
<td>Rain</td>
<td>Cool</td>
<td>Normal</td>
<td>False</td>
<td>P</td>
</tr>
<tr>
<td>Rain</td>
<td>Cool</td>
<td>Normal</td>
<td>True</td>
<td>N</td>
</tr>
<tr>
<td>Overcast</td>
<td>Cool</td>
<td>Normal</td>
<td>True</td>
<td>P</td>
</tr>
<tr>
<td>Sunny</td>
<td>Mild</td>
<td>High</td>
<td>False</td>
<td>N</td>
</tr>
<tr>
<td>Sunny</td>
<td>Cool</td>
<td>Normal</td>
<td>False</td>
<td>P</td>
</tr>
<tr>
<td>Rain</td>
<td>Mild</td>
<td>Normal</td>
<td>False</td>
<td>P</td>
</tr>
<tr>
<td>Sunny</td>
<td>Mild</td>
<td>Normal</td>
<td>True</td>
<td>P</td>
</tr>
<tr>
<td>Overcast</td>
<td>Mild</td>
<td>High</td>
<td>True</td>
<td>P</td>
</tr>
<tr>
<td>Overcast</td>
<td>Hot</td>
<td>Normal</td>
<td>False</td>
<td>P</td>
</tr>
<tr>
<td>Rain</td>
<td>Mild</td>
<td>High</td>
<td>True</td>
<td>N</td>
</tr>
</tbody>
</table>

Q3 a) Discuss Data Warehouse design strategies in detail? [10]

b) Discuss the types of attributes and data visualization for data exploration. [10]
Q4 a) Discuss various OLAP models and their architecture. [10]

b) Find clusters using k-means clustering algorithm, if we have several objects (4 types of medicines) and each object have two attributes or features as shown in table below. The intention is to group these objects into k = 2 group of medicine based on the two features (pH and weight index).

<table>
<thead>
<tr>
<th>Object</th>
<th>Attribute 1 (X) Weight Index</th>
<th>Attribute 2 (Y) pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicine A</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Medicine B</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Medicine C</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Medicine D</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

Q5 a) Discuss the process of extraction, transformation and loading with a neat and labelled diagram. [10]

b) A database has five transactions. Let minimum support = 40% and minimum confidence = 60%
   i) Find all frequent patterns using Apriori Algorithm.
   ii) List strong association rules.

<table>
<thead>
<tr>
<th>Transaction-Id</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1, 3, 4, 6</td>
</tr>
<tr>
<td>B</td>
<td>2, 3, 5, 7</td>
</tr>
<tr>
<td>C</td>
<td>1, 2, 3, 5, 8</td>
</tr>
<tr>
<td>D</td>
<td>2, 5, 9, 10</td>
</tr>
<tr>
<td>E</td>
<td>1, 4</td>
</tr>
</tbody>
</table>

Q6 Write short note on the following (Answer any FOUR) [20]
   i) Applications of Data Mining (minimum two in detail)
   ii) Data pre-processing
   iii) FP Tree
   iv) Updates to dimension tables
   v) Meta data with example