Name: **Enrolment No:**	UPES UNIVERSITY WITH A PURPOSE

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2019

Course: Data Warehouse and Data Mining **Semester: VII**
Program: B.Tech CSE-Cyber Laws **Time 03 hrs.**
Course Code: CSEG 411 **Max. Marks: 100**

Instructions:

SECTION A

S. No.		Marks	CO
Q 1	Why metadata is important? 'How to provide metadata in warehouse?	**04**	**CO1**
Q 2	What is ETL? Explain the steps in ETL.	**04**	**CO2**
Q 3	What is classification in data mining? Explain 2 classification-models with example.	**04**	**CO3**
Q 4	Develop the Apriori algorithm for generating frequent-itemset.	**04**	**CO4**
Q 5	What is Bayes Theorem? Show how it is used for classification.	**04**	**CO5**

SECTION B

Q 1	Identify and describe the phases in the KDD process. How does KDD differ from data mining?	**10**	**CO4**
Q 2	Design a star-schema, snowflake schema and fact constellation schema for the data warehouse that consists of the four dimensions: (Time, Item, Branch and Location). Include the appropriate measures for the schemas.	**10**	**CO2**
Q 3	What is a data warehouse? What is the need for a data warehouse? Also discuss the goals of a data warehouse.	**10**	**CO1**
Q 4	What is hierarchical clustering method? Explain the algorithms for computing distances between clusters. OR How decision-trees are used for classification? Explain decision-tree induction algorithm for classification	**10**	**CO5**

SECTION-C

<table>
<tr>
<td>Q 1</td>
<td>Consider the following transaction database:

<table>
<tr><th>TID</th><th>Items</th></tr>
<tr><td>1</td><td>A,B,C,D</td></tr>
<tr><td>2</td><td>A,B,C,D,E,G</td></tr>
<tr><td>3</td><td>A,C,G,H,K</td></tr>
<tr><td>4</td><td>B,C,D,E,K</td></tr>
<tr><td>5</td><td>D,E,F,H,L</td></tr>
<tr><td>6</td><td>A,B,C,D,L</td></tr>
<tr><td>7</td><td>B,I,E,K,L</td></tr>
<tr><td>8</td><td>A,B,D,E,K</td></tr>
<tr><td>9</td><td>A,E,F,H,L</td></tr>
<tr><td>10</td><td>B,C,D,F</td></tr>
</table>

Apply the Apriori algorithm with minimum support of 30% and minimum confidence of 70% and find all the association rule in the data set.
OR
Given the following transactions, the frequent item set mining should be performed, using the frequent-pattern (FP) growth approach and a minimum support of 3.

<table>
<tr><th>TID</th><th>Items</th></tr>
<tr><td>1</td><td>F,A,C,D,G,I,M,P</td></tr>
<tr><td>2</td><td>A,B,C,F,L,M,O</td></tr>
<tr><td>3</td><td>B,F,H,,J,O,W</td></tr>
<tr><td>4</td><td>B,C,K,S,P</td></tr>
<tr><td>5</td><td>A,F,C,E,L,P,M,N</td></tr>
</table>

Construct the FP-tree corresponding to the set of transactions in an above table. The results should include the set of frequent patterns generated through the different steps in the analysis.</td>
<td>20</td>
<td>CO5</td>
</tr>
<tr>
<td>Q 2</td>
<td>Write short note on following
(a) Web content mining
(b) Comparison between OLAP and OLTP
(c) DMQL
(d) Spatial clustering Algorithm.</td>
<td>20</td>
<td>CO3</td>
</tr>
</table>