

Course Code:		CST411		
Seventh Semester B.E (Computer Science and Engineering) Examination				
DATA WAREHOUSING AND MINING				
Time: 2 Hours]			[Max. Marks: 40	
Instructions to Candidates:				
<div>1. Assume suitable data wherever necessary and clearly state your assumptions.</div> <div>2. Give examples wherever necessary</div> <div>3. Make use of Graph paper</div> <div>4. Write your paper in neat and clean handwriting</div>				
Question		Description of Question	Marks	CO
1	(a)	<p>Bank wants to develop a data warehouse for effective decision-making about their loan schemes. The banks provide loans to customers for various purposes like House Building Loan, Car Loan, Educational Loan, Personal Loan, etc. The whole country is categorized into a number of regions, namely, North, South, East and West. Each region consists of a set of states. Loan is disbursed to customers at interest rates that change from time to time. Also, at any given point of time, the different types of loans have different rates. The data warehouse should record an entry for each disbursement of loan to customer.</p> <p>With respect to the above use case, answer the following questions. Necessary assumptions can be made to support your answer.</p> <div><div>(i) Design a star schema for the data warehouse clearly identifying the fact table(s), dimensional table(s), their attributes and measures along with the primary key and foreign key relationships.</div><div>(ii) Write an SQL query by which you can display region-wise, bank-wise, year-wise total amount of loans disbursed from your schema.</div><div>(iii) List possible schema hierarchies for each dimension.</div></div>	06	CO1
2	(a)	<p>The Employee Details View (EMP_DETAILS_VIEW) provides inner joins between the employees, departments, jobs, locations, countries, and regions tables.</p> <p>EMP_DETAILS_VIEW (EMPLOYEE_ID, JOB_ID, MANAGER_ID, DEPARTMENT_ID, LOCATION_ID, COUNTRY_ID, FIRST_NAME, LAST_NAME, SALARY, COMMISSION_PCT, DEPARTMENT_NAME, JOB_TITLE, CITY, STATE_PROVINCE, COUNTRY_NAME, REGION_NAME)</p> <div><div>(i) What is the output of the following query?</div><div>SELECT decode(grouping(region_name),0,region_name,'GRAND') AS region_name, decode(grouping(country_name),0,country_name,'TOTAL') AS country_name, count(*)</div></div>	06	CO2

		FROM emp_details_view GROUP BY ROLLUP(region_name, country_name); (ii)Write a SQL query that returns the count of employees by region, country, state and department. (iii)Add a ROLLUP clause to the query in (ii) and comment on the results. (iv)Rewrite the query to add calls to the GROUPING function for each level. (v)Limit the results to only display the detail, region_name and grand totals levels.														
3	(a)	Create Index organized table Supplier with attributes Sup_No, Sup_Name, Sup_Email, Sup_Type, Sup_phone , Sup_address and Sup_Remarks in tablespace 'tsa'. Sup_phone, Sup_address and Sup_Remarks should be in Overflow block with Pctthreshold as 20.	04	CO2												
	(b)	Implement Reference Partitioning on data warehouse for the following scenario: <ul style="list-style-type: none">• Create parent table Invoice with the attributes Inv_id, Inv_date, cust_id, Inv_Amount• Perform Range partitioning on Inv_date. Take Range of 04 Months• Create child table Inv_Line_Items with attributes Inv_id, Item_id, Item_Name, price and quantity.• Perform Reference partitioning on child table.	03	CO2												
4	(a)	The choice of beverage (coffee, tea, or milk) of a random sample of people is shown in the table. At significant level $SL = 0.10$, what can we conclude about the relationship between choice of beverage and the region they live in? [Given : for $SL = 0.10$ and degree of freedom = 2, the critical value is 4.605] <table><tr><td></td><td>Coffee</td><td>Tea</td><td>Milk</td></tr><tr><td>Region-1</td><td>58</td><td>81</td><td>61</td></tr><tr><td>Region-2</td><td>42</td><td>19</td><td>39</td></tr></table>		Coffee	Tea	Milk	Region-1	58	81	61	Region-2	42	19	39	04	CO3
	Coffee	Tea	Milk													
Region-1	58	81	61													
Region-2	42	19	39													
4	(b)	Suppose two stocks ABC Corp. and PQR Ltd. have the following values in one week: (3, 6), (4, 9), (6, 11), (5, 12), (7, 15) If the stocks are affected by the same industry trends, using covariance comment on the whether the stocks will rise or fall together.	03	CO3												

5	(a)	<p>Generate the Frequent Patterns for the following dataset using FP-growth algorithm. Show the conditional pattern base, conditional FP-tree, and the frequent patterns generated. Given min_sup = 2.</p> <table><tr><td>TID</td><td>Items</td></tr><tr><td>1</td><td>{a,b}</td></tr><tr><td>2</td><td>{b,c,d}</td></tr><tr><td>3</td><td>{a,c,d,e}</td></tr><tr><td>4</td><td>{a,d,e}</td></tr><tr><td>5</td><td>{a,b,c}</td></tr><tr><td>6</td><td>{a,b,c,d}</td></tr><tr><td>7</td><td>{a}</td></tr><tr><td>8</td><td>{a,b,c}</td></tr><tr><td>9</td><td>{a,b,d}</td></tr><tr><td>10</td><td>{b,c,e}</td></tr></table>	TID	Items	1	{a,b}	2	{b,c,d}	3	{a,c,d,e}	4	{a,d,e}	5	{a,b,c}	6	{a,b,c,d}	7	{a}	8	{a,b,c}	9	{a,b,d}	10	{b,c,e}	07	CO4														
TID	Items																																							
1	{a,b}																																							
2	{b,c,d}																																							
3	{a,c,d,e}																																							
4	{a,d,e}																																							
5	{a,b,c}																																							
6	{a,b,c,d}																																							
7	{a}																																							
8	{a,b,c}																																							
9	{a,b,d}																																							
10	{b,c,e}																																							
6	(a)	<p>Use single, complete, and average link agglomerative clustering to group the data described by the following distance matrix. Produce the dendrograms.</p> <table><tr><td></td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>1</td><td>0</td><td></td><td></td><td></td><td></td></tr><tr><td>2</td><td>9</td><td>0</td><td></td><td></td><td></td></tr><tr><td>3</td><td>3</td><td>7</td><td>0</td><td></td><td></td></tr><tr><td>4</td><td>6</td><td>5</td><td>9</td><td>0</td><td></td></tr><tr><td>5</td><td>11</td><td>10</td><td>2</td><td>8</td><td>0</td></tr></table>		1	2	3	4	5	1	0					2	9	0				3	3	7	0			4	6	5	9	0		5	11	10	2	8	0	07	CO4
	1	2	3	4	5																																			
1	0																																							
2	9	0																																						
3	3	7	0																																					
4	6	5	9	0																																				
5	11	10	2	8	0																																			