B.E. (Information Technology) Seventh Semester (C.B.S.)
Data Warehousing & Mining

_____

Notes :
1. All questions carry marks as indicated.
2. Solve Question 1 OR Questions No. 2.
3. Solve Question 3 OR Questions No. 4.
4. Solve Question 5 OR Questions No. 6.
5. Solve Question 7 OR Questions No. 8.
6. Solve Question 9 OR Questions No. 10.
7. Solve Question 11 OR Questions No. 12.
8. Due credit will be given to neatness and adequate dimensions.
9. Assume suitable data whenever necessary.

1. a) Describe the failure of past decision support system. **6**

b) What do you mean by strategic information? Explain the factor that escalated the need for strategic information. **7**

OR

2. a) Describe Data Ware House Architecture in detail. **7**

b) What is data mart? How it differs from data warehouse? **6**

3. a) What is data reduction? Explain different methods of data reduction. **6**

b) Summarize the data preprocessing steps in brief. **7**

OR

4. a) Enlist the measures of descriptive data summarization. **7**

b) Explain Min-max normalization and Z-Score normalization. **6**

5. a) Describe STAR and SNOW FLAKE schema with example. **8**

b) Define Data cube and explain OLAP operations on data cubes. **6**

OR

6. a) What are different types of OLAP models? **6**

b) What is the difference between OLAP and OLTP? **4**

c) What is the need for multi-dimensional analysis? **4**

7. a) Draw and explain knowledge Discovery Process. **7**

   b) Describe the classification of Data Mining Systems. **6**

OR

8. a) Draw and explain data mining architecture. **7**

   b) Write the applications of Data Mining. **6**

9. a) Define the following terms with examples. **6**
     i) Frequent itenset   ii) Closed intenset

   b) What is Market Basket Analysis? Explain in detail with example. **8**

OR

10. A database has five transactions Let min-sup=60% and min-conf = 80%. **14**

TID	items-bough
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y}
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I, E}

Find all frequent intensets using Apriori and FP-growth respectively compare the efficiency of two mining processes.

11. a) Write in details about BI framework 2020 with neat diagram. **7**

   b) What is BI? Explain important factors of BI. **6**

OR

12. a) What is Big-data? What are the challenges Big-data? **6**

   b) Describe the architectural components of Hadoop ecosystem. **7**

****