END TERM EXAMINATION

FOURTH SEMESTER [MCA] MAY-2010

Paper Code: MCA 202   Subject: Data Warehousing and Data Mining
Paper ID: 44202

Time : 3 Hours   Maximum Marks : 60

Note: Question 1 is compulsory. Attempt one question from each unit.

Q1. (a) What is an information crisis?   (2 x 10=20)
 (b) Explain different types of data used in data warehouse.
 (c) What is a data mart? Cive example.
 (d) What is a factless fact table? Explain with example.
 (e) How does a snowflake schema differ from STAR Schema?
 (f) What are junk dimensions? Are they necessary in a data warehouse?
 (g) Discuss reasons why feeding data into the OLAP system directly from the source operational system is not recommended.
 (h) What are various factors for consideration in OLAP administration?
 (i) What are the criteria for evaluating data mining tools?
 (j) How is data mining the primary step in the process of knowledge discovery?

Unit-I

Q2. (a) You are senior analyst in the IT department of a company manufacturing automobile parts. The marketing VP is complaining about the poor response by IT in providing strategic information. Draft a proposal to him explaining the reason for the problem and why a data warehouse would be the only viable solution.   (8)
 (b) What are advantages and disadvantages of using top- down approach for building a data warehouse?   (2)

Q3. (a) What is information package diagram (IPD)? How does it help in dimensional analysis ? Make an IPD of sale anaysis System.   (7)
 (b) What do you mean by information system? How is different from operation System?   (3)

Unit-II

Q4. (a) What is star schema? What are its component tabes? Expain by considering suitable example of a supermarket chain.   (7)
 (b) Why is the entity- relationship modeling technique not suitable for a data warehouse? How is dimensional modeling different?   (3)

Q5. (a) A sale organization builds its data warehouse. It is hnown that not all products are sold at each outlet every day. Show how the use of aggeregatin can speedup analysis.   (5)
 (b) Differentiate between slowly changing and rapidly changing dimensions. (5)

Unit-III

Q6.   (a) You are asked to form to evaluate the MOLAP and ROLAP models and make your recommendations for large manufacturer of heavy chemicals. Describe the criteria your team will use to make the evaluation and selection.   (7)

           (b) What are hypercubes? How do they apply in an OLAP system?   (3)

Q7.   (a) Pick any five of codd's guidelines for OLAP, give reason why the selected guidelines are important for OLAP.   (5)

           (b) Exlain different OLAP operations with suitable example.   (5)

Unit-IV

Q8.   (a) What is data mining? Compare various data mining techniques.   (5)

           (b) Explain knowledge variows discovery process in details.   (5)

Q9.   (a) Do neural network and genetic algorithm have anything common? Point out few differences?   (5)

           (b) How is data mining different from OLAP? Explain briefly.   (5)