

END TERM EXAMINATION

FOURTH SEMESTER [MCA]MAY JUNE 2012

Paper Code: MCA 204

Subject: Data Warehousing & Data Mining

Time:3 Hours

Maximum Marks:60

Note: Attempt five question including Q.no.1 which is compulsory. Select one Question from each unit.

Question 1 :-

- (a)What is strategic information?
- (b) What is data mart? When is it appropriate?
- (c)Data warehousing and data mining creates intelligence in business. Justify.
- (d)What are the advantages of snowflake schema over star schema?
- (e)What are hypercubes?.
- (f)Discuss reasons why feeding data into the OLAP system directly from the source operational system is not recommended?
- (g)Is the data warehouse a pre-requisite for data mining? Why/Why not?
- (h)Define support and confidence for an association rule.
- (i)Explain correlation analysis for handling redundancy.
- (j)Compute the Euclidean and Manhattan distance between the two objects represented by following tuples (1,6,2,5,3) and (3,5,2,6,6). **(2*10=20)**

UNIT-I

Q2(a)Describe five differences between operational system and informational systems. **(5)**

(b)Explain data granularity and how it is applicable to the data warehouse. **(5)**

Q3(a)What a information package diagram(IPD)? How it helps in dimensional analysis? Make an IPD of Hotel occupancy system.. **(7)**

(b)What are the advantages and disadvantages of using bottom-up approach for building a data warehouse? (3)

UNIT II

Q4(a)What is dimensional modeling? How it is different from E-R modeling? (3)

(b)What is STAR schema? Explain by taking example. A dimension table is wide and fact table is deep. Justify. (7)

Q5(a)Describe slowly changing dimensions. What are the three types of change? Explain each type briefly. (5)

(b)Draw and explain architecture of MOLAP. What are the advantages of MOLAP over ROLAP? (5)

UNIT-III

Q6(a)How is data mining different from OLAP? What are the advantages of data mining? Explain briefly.. (5)

(b)Explain knowledge discovery Process(KDD) in details.. (5)

Q7(a)What is data cleaning? Explain various smoothing techniques of noisy data. (5)

(b)List and describe the five primitives for specifying a data mining task. (5)

UNIT-IV

Q8(a)You are the marketing manager of LG electronics and would like to characterize the buying habits of customers who PURCHASE items in North India [city<state<region<country] and priced not less than Rs 5000 w.r.t. CUSTOMER's age, type of ITEM purchased, the place in which ITEM was made. For each characteristic discovered, you would like to know the % of customers with 5% noise threshold having that characteristic. The result should be in the tabular form. Write DMQL for it. (5)

(b) What is classification? Describe decision tree technique of classification with example. (5)

Q9(a) The following six objects each with two attributes are to be clustered
A1(4,6) A2(2,6) A3(9,3) A4(6,9) A5(7,5) A6(5,7) :-

- (i) Show the distance matrix for six objects using the Manhattan distance.
- (ii) Using the divisive methods determine the two objects that should form the basis for splitting the above dataset.
- (iii) Now split the dataset using the two objects identified in part(ii) using the k-means method. **(5)**
- (b) Briefly explain different data mining applications. **(5)**