

Code: 06MC403

M.C.A IV Semester Supplementary Examinations, June 2012

**DATA WAREHOUSING AND MINING**

(For 2006, 2007 & 2008 admitted batches only)

Time: 3 hours

Max Marks: 60

Answer any FIVE questions

All questions carry equal marks

\*\*\*\*\*

- 1 (a) Explain major issues in mining and data mining primitives.  
(b) What are data stores, data repositories, flat files and explain functionalities.
- 2 Explain designing dimension tables like snow flake and star schema.
- 3 Consider a fact table F (year, location, product, sale amount), let year take 15 values, location 100 values, product 150 values, consider a completely materialized full data cube with all grid cells. Consider a fact table row (2011, Hyd, bitter gourd cake, 500), suppose this row had errors, state what is the most efficient way of updating the complete cube in each of the following cases.  
(a) Actual row is (2011, Hyd, bitter guard cake,750)  
(b) Actual row.
- 4 (a) Attribute-oriented induction generates one or a set of generalized descriptions. How can there descriptions be visualized?  
(b) Discuss about the methods of attribute relevance analysis.
- 5 What is decision tree? What are the advantages and disadvantages of decision tree classification for the given data set create a decision tree? Explain about the knowledge obtained from it outlook, temp(F) Humidity windy class  
Sunny,79,90,true play  
Sunny, 76,70, flase play  
Sunny, 79, 75, true, no, play  
Sunny, 60, 90, true, no, play.  
Overcast, 88, 88, flase, no, play.  
Overcast, 63, 75, true, play  
Rain, 78,60, flase, play  
Rain, 68, 60, true, play.

Contd. in Page 2

Code: 06MC403

- 6 (a) Explain Apriori algorithm with an example.  
(b) Explain market basket analysis and its relevance to association rule.
- 7 Compare and contrast hierarchical and density-based clustering methods illustrate with examples.
- 8 Discuss about:  
(a) Textual databases  
(b) Spatial databases  
(c) Time-series databases  
(d) Web database.

\*\*\*\*\*

JNTUWORLD