DHARMSINH DESAI UNIVERSITY, NADIAD
FACULTY OF TECHNOLOGY
B.TECH. SEMESTER V [I.T]
SUBJECT: (IT-704) Data Analysis & Information Extraction

Examination	**: First Sessional**	**Seat No.**	**: _____**
Date	**: 01/08/2014**	**Day**	**: Friday**
Time	**: 1:00 to 2:15**	**Max. Marks**	**: 36**

INSTRUCTIONS:
1. Figures to the right indicate maximum marks for that question.
2. The symbols used carry their usual meanings.
3. Assume suitable data, if required & mention them clearly.
4. Draw neat sketches wherever necessary.

Q.1 Do as directed.
(a) Define data mining. **[2]**
(b) Why do we require separate data warehouse? **[2]**
(c) Differentiate between data warehouse and data mart. **[2]**
(d) Calculate number of cuboids for 8 dimensions and each dimension has 5 levels. **[2]**
(e) Redundancy is an important issue in D.W. How can you deal with it? **[2]**
(f) Number of users in OLTP is smaller than OLAP. True /False. Justify your answer. **[2]**

Q.2 Attempt ***Any Two*** from the following questions. **[12]**
(a) Explain KDD process with proper diagram. **[6]**
(b) Explain three Tier data warehouse architecture with proper figure. **[6]**
(c) Explain dimensionality reduction methods for data reduction. **[6]**

Q.3 (a) Identify dimensions and measures. Draw star schema diagram and 3-D CUBE for university management system. **[6]**
(b) Suppose data analysis for age attribute as per given below. **[6]**
13,15,16,16,19,20,20,21,22,22,25,25,25,25,30,33,33,35,35,35,35,36,40,45,46,52,70
i) Use min-max normalization to transform the value 35 for age onto the range [0.0, 1.0].
ii) Use Z-score normalization to transform the value 35 for age where standard deviation for age is 12.94 years.
iii) Comment on which method you would prefer to use for the given data giving reasons as to why.

OR

Q.3 (a) Identify dimensions and measures. Draw star schema diagram and 3-D CUBE for Hospital management system. **[6]**
(b) Apply segmentation by natural partitioning using 3-4-5 rule for given data and also draw Concept Hierarchy. For All Electronics company profits at different branches cover wide range -$360,976.00 to $4,700,896.50. Suppose that data within 5^{th} percentile and 95^{th} percentile are between -$159,876 and $1,838,761. **[6]**