

Seat No.: _____

Enrolment No. _____

GUJARAT TECHNOLOGICAL UNIVERSITY
MCA SEMESTER—IV –EXAMINATION –SUMMER-2016

Subject code: 640005

Date: 04-06-2016

Subject Name: Data Warehousing & Data Mining

Time: 10.30a.m. To 01.00p.m.

Total Marks: 70

Instructions:

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.

- Q.1** (a) What is Data Warehouse? Explain the following Terms for Data Warehouse **07**
1. Subject-oriented.
2. Integrated.
3. Time-variant.
4. Nonvolatile.
- (b) Discuss the Differences between Operational Database Systems and Data Warehouses. **07**
- Q.2** (a) What is top-down approach? Discuss warehouse design process steps. **07**
(b) Discuss following models for Data Warehouse **07**
1. Enterprise warehouse.
2. Data mart
3. Virtual warehouse.

OR

- (b) Discuss Range, Quartiles, Outliers, and Boxplots With Proper example. **07**
- Q.3** (a) Explain the Concept of Relational Databases with Proper Example. **07**
(b) How is class comparison performed? Give the Details of general procedure of class comparison. **07**

OR

- Q.3** (a) What is Data Mining? Discuss the major components of Data Mining. **07**
(b) Discuss Full materialization and Partial materialization with difference. **07**
- Q.4** (a) Discuss Frequent Pattern Mining: A Road Map. **07**
(b) What Is Classification? What Is Prediction? Discuss the Preparing the Data for Classification and Prediction. **07**

OR

- Q.4** (a) Explain Apriori Algorithm with example. **07**
(b) Give the Short note on: - Decision Tree Induction. **07**
- Q.5** (a) What is Cluster Analysis? Discuss Classical Partitioning Methods: k-Means and k-Medoids. **07**
(b) Discuss Data Mining for Financial Data Analysis. **07**

OR

- Q.5** (a) Explain : Distance-Based Outlier Detection **07**
(b) Discuss Data Mining for Biological Data Analysis with DNA sequence and corresponding amino acid sequence. **07**
