No. of Printed Pages : 7 **MBMI-011**

MBA – INFORMATION TECHNOLOGY MANAGEMENT (MBAITM)

Term-End Examination

00284 **December, 2014**

MBMI-011 : DATA WAREHOUSING AND DATA MINING

Time : 3 hours *Maximum Marks : 100*

Note :

(i) *Section I is* ***compulsory*** *and carries 30 marks*

(ii) *Section II carries 70 marks. Answer any* ***five*** *questions.*

(iii) *Assume suitable data/assumptions wherever required.*

(iv) *Draw suitable sketches wherever required.*

(v) *Italicized figures to the right indicate maximum marks.*

SECTION I

1. Case study :

NTI is one of the premier private education institutions. The institute offers various courses in Engineering at capacity of post-graduate and undergraduate level. The courses are :

- B.Tech. (Bachelor of Technology)/ B.E. (Bachelor of Engineering)
- M.S. (Master of Sciences)
- M.Tech. (Master of Technology)

All courses are fully residential programs with duration of 2 years (four semesters). The campus is equipped with the following facilities :

- International standard Cafeteria
- Hostel facility for both boys and girls
- Wi-Fi connection
- State-of-the-art infrastructure
- Dedicated placement department
- Experienced faculty

The selection process for admission requires an online test. Students fill the following online application with the details like,

- Name, Address, E-mail, Date-of-birth, Gender and Blood group.
- Programs choice
- Payment details (DD no. Date, Bank name)
- Details of the previous education (SSC to BE/B.Tech/MCA/MSC)
- Program dates
- Previous academic scores
- Academic achievements

Currently the institute maintains the following operational databases at departmental level :

- Online admission test system
- Employee payroll
- Examination system
- Student in/out register system
- Hostel administration

Grade System

- The Institute follows G.P.A. system in each semester. Maximum grade is set to 4·0 G.P.A.
- The G.P.A. is calculated based on the credits of the course and grade obtained in the examination.

Problems :

The current system has the following problems for effective decision-making :

- Unable to take consolidated decisions about issues like grace marks, etc.
- Multiple report generation for single dimensional analysis
- Poor data analytical techniques
- Un-skilled data managers
- No consolidated data view

To solve the above problem, an analyst has suggested data warehouse implementation. Answer the following questions related to data warehouse implementation for the current system :

Questions :

- Draw suitable architecture for solving the data integration problem. Explain its advantages. *10*
- Identify one subject area for analysis. *3*
- Draw the Information Package Diagrams for requirements. *5*
- Design Dimensions with Type-1, Type-2 and Type-3 Slowly Changing dimensions. *7*
- Draw the Star Schema Diagram with necessary data integrity keys. *5*

SECTION II

2. (a) For an airlines company, how can strategic information increase the number of frequent flyers ? Discuss giving specific details. *3+3*

(b) What is Strategic Information ? Explain the factors relating to inability in procuring strategic information. *3+5*

3. (a) Data warehouse is subject-oriented. What would be the major critical business subjects for the following companies :

(i) An international manufacturing company *3*

(ii) A local community bank *3*

(iii) A domestic hotel chain *3*

(b) Every data structure in data warehouse contains time element. Why ? Write reasons. *5*

4. (a) As a data designer of data warehouse for an international bank, consider the possible types of snapshot and transaction tables. Complete the design with one set of snapshot and transaction tables. *8*

(b) You are the Vice President of Marketing for a nation-wide appliance manufacturer with three production plants. Describe any three different ways you will tend to analyze your sales. What are the business dimensions for your analysis ? *3+3*

5. Your project team has decided to use the system logs for capturing the updates from source operational systems. You have to extract data for the incremental loads from four operational systems all running on relational databases. These are four types of sales applications. You need data to update the data in the data warehouse.

   (a) Make assumptions and describe the detailed data extraction process and its architecture. *9*

   (b) What are the most commonly occurring data integration problems ? Explain them briefly. *5*

6. (a) For each one of the following data mining functions (i) to (iv), suggest suitable algorithm and write two application examples for each function :

   (i) Classification *3*

   (ii) Clustering *3*

   (iii) Modelling *3*

   (iv) Sequential Patterns *3*

   (b) What is Click-Stream Analytics ? *2*

7. (a) What is Data mining ? Explain the relationship between data mining and data warehouse. *3+4*

   (b) Is Data mining useful only to identify the business opportunities ? Comment on it. *7*

8. Explain the following Data mining techniques briefly with at least one example for each :

(a) Decision Trees *3*

(b) Case-Based reasons *3*

(c) Neural Computations *4*

(d) Statistical Methods *4*