LINCOLN UNIVERSITY COLLEGE

# EXAMINATION PAPER

| | | |
|---|---|---|
| FACULTY | : | COMPUTER SCIENCE AND MULTIMEDIA |
| COURSE | : | MASTER OF COMPUTER SCIENCE |
| YEAR/ SEMESTER | : | FIRST YEAR / SEMESTER TWO |
| MODULE TITLE | : | DATA WAREHOUSING & BIG DATA |
| CODE | : | DWB 121 |
| TIME ALLOWED | : | 3 HOURS |

## Instruction to candidates

1. This question paper has THREE (3) Sections.
2. Answer ALL questions in Section A, VSAQ.
3. Answer 7 questions out of 9 in Section B, SAQ.
4. Answer 2 questions out of 3 in Section C, LAQ.
5. No scripts or answer sheets are to be taken out of the Examination Hall.

***Do not open this question paper until instructed***

*(Candidates are required to give their answers in their own words as far as practicable)*

## SECTION A

**Very Short Answer Questions**

**Attempt all questions** [7×2=14]

1. What is data mining?
2. Write about the support and confidence in mining.
3. What is the web mining?
4. Define ODS.
5. Write about data cube.
6. State the qualities of good clusters.
7. Mention the issues during text representation.

## SECTION B

**Short Answer Questions**

**Attempt any seven (7) questions out of nine (9) questions[7×8=56]**

1. State the major challenges of mining a huge amount of data (e.g., billions of tuples) in comparison with mining a small amount of data (e.g., data set of a few hundred tuple).
2. "A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision making process". Explain.
3. Discuss how we can relate online airline ticket booking as OLTP.
4. Can we generate equivalent SQL for OLAP queries? Explain with example.
5. "Stock market prediction is an important application of time- series analysis". Describe with reasons.
6. Explain the Apriori Algorithm with its steps.
7. Show the difference between classification and clustering.
8. Briefly explain the different types of web mining.
9. Discuss the Loops in Data warehousing.



## SECTION C

**Long Answer Questions**

**Attempt any two (2) questions out of three (3) questions [2×15=30]**

1. Outliers are often discarded as noise. However, one person's garbage could be another's treasure. For example, exceptions in credit card transactions can help us detect the fraudulent use of credit cards. Using fraudulence detection as an example, propose two methods that can be used to detect outliers and discuss which one is more reliable.
2. Explain the important features of SSIS. Discuss the disadvantages of SSIS. Why do you use Execute SQL Task in SSIS?[6+5+4]
3. The following table consists of training data. Construct a Decision Tree based on this data, using the basic algorithm for decision tree induction. Classify the records by the "Status" attribute. Write down the rules that can be generated from the obtained decision tree.

| Department | Age range | Salary Class | Status |
|---|---|---|---|
| Sales | Middle-aged | High | Senior |
| Sales | Young | Low | Junior |
| Sales | Middle-aged | Low | Junior |
| System | Young | High | Junior |
| System | Middle-aged | High | Senior |
| System | Young | High | Junior |
| System | Senior | High | Senior |
| Marketing | Middle-aged | High | Senior |
| Marketing | Middle-aged | Average | Senior |
| Secretary | Senior | Average | Senior |
| Secretary | Young | Low | Junior |

****BEST OF LUCK****