(Knowledge for Development)

KIBABII UNIVERSITY

UNIVERSITY EXAMINATIONS
2019/2020 ACADEMIC YEAR

SPECIAL/SUPPLEMENTARY EXAMINATIONS
YEAR THREE SEMESTER TWO EXAMINATIONS

FOR THE DEGREE OF
BACHELOR OF SCIENCE COMPUTER SCIENCE

COURSE CODE: CSC 367E

COURSE TITLE: DATA WAREHOUSING & MINING

DATE: 02/02/2021   TIME: 11.00 A.M – 01.00 P.M

INSTRUCTIONS TO CANDIDATES

ANSWER QUESTIONS ONE AND ANY OTHER TWO.

QUESTION ONE (COMPULSORY) [30 MARKS]

a) Define Data mining **[2 Marks]**

b) State why it is important to separate Data Warehouse from operational database.

[5 Marks]

c) What are the main differences between an operational database and a data warehouse?

[4 marks]

d) Explain the four functions performed by OLAP **[4 Marks]**

e) Evaluate the three Schemas that a data warehouse system can implement? **[6 Marks]**

f) There are several datamining techniques, discuss five of these techniques **[5 Marks]**

QUESTION TWO [20 MARKS]

Malaria is a mosquito-borne infectious disease affecting humans and other animals caused by parasitic protozoans. Malaria is spread by mosquitos that bite at night. Malaria is prevalent in Western part of Kenya, this is especially true during rainy seasons. Different areas in Western Kenya have different times when Malaria is prevalent and different people are affected differently by Malaria.

Health ministries in western region have teamed up and decided to carry out some research. They are randomly choosing people and testing them of Malaria (some will have Malaria and some would not have Malaria). Every day a record containing the following data is stored on the person:

- The date
- Percentage of time the person has been outside the net the previous night
- Malaria infection level in the person

A data warehouse is being setup for the purpose of analyzing this data. Furthermore, the ministry of Agriculture sends daily data containing area and the degree to which mosquitos are spreading (zero, low, medium and high).

The researchers have various questions they want to have answered using the data warehouse. For instance:

- People living in certain areas. For each area and month, how many people have low/medium/high malaria infection level
- How many persons do not suffer or suffer to a low, medium or high degree Malaria per day.

a) Draw an architecture diagram for the above described case according to the abstract data warehousing architecture [3 Marks]

b) Give the ER-model of the star schema for the above described data warehouse. Assume people do not leave the area they live in. Furthermore, give a table design for your star schema, i.e., give an overview of the necessary tables with their attributes. [10 Marks]

c) Discuss what need to be changed to the star schema if you assume that people do leave the area they live in for certain days or weeks. [4 Marks]

d) It was assumed originally that they there would be complete information from all people sampled, but this is not the case, occasionally part of the data is lost in transmission whereby data for a couple of days may be lost. Explain how you will handle this data quality problem. [3 Marks]

QUESTION THREE [20 MARKS]

a) What are the differences between the Data Mart Bus Architecture according to Kimball and the Hub and Spoke Architecture according to Inmon? For what context would you choose which architecture? [10 marks]

b) What is meant by a Web-enabled data warehouse? And describe three of its functional features. [10 marks]

QUESTION FOUR [20 MARKS]

a) Define data mining. [2 marks]

b) With the help of examples, explain some operations that can be performed on a data cube. [10 marks]

c) Differentiate between OLAP and OLTP? [4 marks]

d) Compare snowflake to star schemas as used in data warehousing [4 marks]

QUESTION FIVE [20 Marks]

a) With an aid of a diagram explain the Data Mining Architecture. [10 marks]

b) In data mining what is correlation analysis used for? [2 marks]

c) Using examples explain four data mining application areas [8 marks]