



KIBABII UNIVERSITY COLLEGE

*(A Constituent College of Masinde Muliro University of Science
Technology)*

P.O. Box 1699-50200 Bungoma, Kenya

Tel. 020-2028660/0708-085934/0734-831729

E-mail: enquiries@kibabiiuniversity.ac.ke

**UNIVERSITY EXAMINATIONS
2013/2014 ACADEMIC YEAR**

THIRD YEAR SECOND SEMESTER EXAMINATIONS

**FOR THE DEGREE OF BACHELOR
OF INFORMATION TECHNOLOGY**

COURSE CODE: BIT 324/CSC 367E

COURSE TITLE: DATA WAREHOUSING AND MINING

DATE: 17TH APRIL, 2014

TIME: 9:00A.M. – 12 NOON.

INSTRUCTIONS TO CANDIDATES

- Answer question ONE and any other two questions
- Time Allowed: 3 hours

Section A: Answer ALL questions in this section

Question One (30 marks)

- a) Distinguish between the following data warehouse concepts. [3 marks]
- i) Data warehouse and data mart
 - ii) Base cuboid and apex cuboid
 - iii) Dimension and fact
- b) The data warehouse architecture is the proper arrangement of the software and hardware components. With the aid of a diagram explain the building blocks or components of data warehouse that suit the requirements of any organization for maximum benefit. [10 marks]
- c) Metadata in a data warehouse fall into three major categories. Briefly explain the purpose of each type.
- i) Operational Metadata [2 marks]
 - ii) Extraction and Transformation Metadata [2 marks]
 - iii) End User Metadata [2 marks]
- d) Name any six different methods for information delivery. [3 marks]
- e) **“A Data Warehouse is a subject oriented, integrated, nonvolatile, and time variant collection of data in support of management’s decisions.”** The four keywords, *subject-oriented*, *integrated*, *time-variant*, and *nonvolatile*, are defining features that distinguish data warehouses from other data repository systems, such as relational database systems, transaction processing systems, and file systems. Briefly explain these defining features with regard to data, nature of data, and usage. [8 marks]

Section B: Answer any TWO questions in this section

Question Two (20 marks)

- a) What is data visualization? Briefly describe the three remarkable advanced data visualization techniques. [4 marks]
- b) The major task of on-line operational database systems is to perform on-line transaction and query processing. Data warehouse systems, serve users or knowledge workers in the role of data analysis and decision making. Such systems can organize and present data in various formats in order to accommodate the diverse needs of the different users. Explain the difference between OLTP and OLAP. [10 marks]
- c) Data warehouses generalize and consolidate data in multidimensional space. The most popular data model for a data warehouse is a multidimensional model. This data model can exist in three main forms. Describe the three forms of the multidimensional data model in detail. [6 marks]

Question Three (20 marks)

Consider the following business scenario. The table below shows a 3-D view of sales data for *AllElectronics*, according to the dimensions *time*, *item*, and *location*. The measure displayed is *dollars_sold* (in thousands).

Time (quarter)	Location = "Chicago"				Location = "New York"				Location = "Toronto"				Location = "Vancouver"			
	Home Ent.	Comp.	Phone	Security	Home ent.	Comp.	Phone	Security	Home ent.	Comp.	Phone	Sec	Home ent.	Comp.	Phone	Sec
Q1	854	882	89	623	1087	968	38	872	818	746	43	591	605	825	14	400
Q2	943	890	64	698	1130	1024	41	925	894	769	52	682	680	952	31	512
Q3	1032	924	59	789	1034	1048	45	1002	940	795	58	728	812	1023	30	501
Q4	1129	992	63	870	1142	1091	54	984	978	864	59	784	952	1038	38	580

Just as relational query languages like SQL can be used to specify relational queries, a data mining query language can be used to specify data mining tasks. Define a multidimensional schema for these data using SQL-based data mining query language (DMQL) using two language primitives, one for cube definition and one for dimension definition syntax for each of the following:

- i) Star schema definition [6 marks]
- ii) Snowflake schema definition. [6 marks]
- iii) Fact constellation schema definition [8 marks]

Question Four (20 marks)

- a) Define data mining. [1 marks]
- b) Name the major phases of a data mining operation. Out of these phases, pick two and describe the types of activities in these two phases. [5 marks]
- c) How is data mining different from OLAP? Explain briefly. [8 marks]
- d) Is the data warehouse a prerequisite for data mining? Does the data warehouse help data mining? If so, in what ways? [3 marks]
- e) Briefly describe the cluster detection data mining technique. [3 marks]