



**MASINDE MULIRO UNIVERSITY OF
SCIENCE AND TECHNOLOGY**

(MMUST)

MAIN CAMPUS

UNIVERSITY EXAMINATIONS

2021/2022 ACADEMIC YEAR

**THIRD YEAR SECOND SEMESTER EXAMINATIONS
FOR THE DEGREE**

OF

BACHELOR OF SCIENCE IN COMPUTER SCIENCE

COURSE CODE: BCS 368

COURSE TITLE: DATA WAREHOUSING & MINING

DATE: 22/04/2022

TIME: 8:00-10:00AM

INSTRUCTIONS TO CANDIDATES

Question ONE (1) is compulsory

Answer TWO (2) questions

TIME: 2 Hours

MMUST observes ZERO tolerance to examination cheating

SECTION A: (Answer all the Questions in this Section)

QUESTION ONE (30 Marks)

- a) Explain any five trends in data warehousing? (5 Marks)
- b) State four major data cleaning tasks? (4 Marks)
- c) Compare OLTP and OLAP Systems? (4 Marks)
- d) MMUST is implementing a data warehouse to keep track of the student population and bio information from its inception. The amount of data is continually expanding and therefore a three tier data warehouse is required to take care of this repository. Using a well-illustrated diagram, explain the data warehouse architecture that you would present to university to help with the data repository implementation? (6 Marks)
- e) List two techniques of data mining? (2 Marks)
- f) Differentiate between a regular user and a power user? (2 Marks)
- g) State five reasons why we should use a data warehouse (5 Marks)
- h) State any two categories of Metadata? (2 Marks)

SECTION B: (Answer any 2 Questions in this Section)

QUESTION TWO (20 Marks)

- i.) Explain any five challenges of implementing a data mine? (10 Marks)
- ii.) State four application of data mining (4 Marks)
- iv.) “Snowflaking” is a method of normalizing the dimension tables in a STAR schema. State any four methods to consider before normalizing the tables? (4 Marks)
- v.) State two advantages of “snowflaking” in database normalization? (2 Marks)

- ii.) State three advantages of Star Schema technique of data mining? (3 Marks)
- iii.) Jumia is an online e-commerce site that deals with selling different types of products in the market. Most of the clients or customers purchase items which are delivered via courier to their preferred pickup points. They have therefore acquired a three-layer data warehouse to help keep track of its online operations with yearly backups being loaded to the data warehouse. You have been employed by Jumia as their chief data analyst. Explain the data transformation tasks that you will perform in order to load their data into the newly implemented data warehouse? (10 Marks)
- iv.) A lot of data is available today. State two major sources of data? (2 Marks)

QUESTION FOUR (20 Marks)

- i.) Explain the meaning of data Quality? (2 Marks)
- ii.) A lot of data is pumped in the internet. Most of the data are unstructured in nature therefore making it very difficult to analyze such types of data. The other section of the data posted to the internet by various users also include the structured data which is organized in the relational database structure. You have been requested to discuss what differences occur in the data posted in the internet as either structured, unstructured and semi-structured data in terms of the quality of data they contain. Explain the characteristics that will qualify the data to be either of quality or not? (5 Marks)
- iii.) The Quality of data in a data warehouse is affected by various issues. State five problems that affect the quality of data. (5 Marks)
- iv.) Differentiate between data integration and data transformation (2 Marks)
- v) Explain three data integration issues? (6 Marks)

QUESTION FIVE (20 Marks)

- i.) In which architectural layer does OLAP fit? (1 Mark)
- ii.) What are the benefits of OLAP? (4 Marks)
- iii.) List five major functions and services in the data storage area? (5 Marks)
- iv.) State five functions of queries and reports in a data warehouse? (5 Marks)

v.) Explain three major reasons why metadata is vital for end-users. (3 Marks)

vi.) Differentiate between Spatial OLAP and Mobile OLAP? (2 Marks)