MERU UNIVERSITY OF SCIENCE AND TECHNOLOGY

P.O. Box 972-60200 – Meru-Kenya.
Tel: +254(0) 799 529 958, +254(0) 799 529 959, +254 (0)712 524 293
Website: www.must.ac.ke Email: info@mucst.ac.ke

UNIVERSITY EXAMINATIONS 2023/2024

FOURTH YEAR, SECOND SEMESTER EXAMINATION FOR THE DEGREE OF BACHELOR OF BUSINESS INFORMATION TECHNOLOGY, BACHELOR OF INFORMATION TECHNOLOGY AND BACHELOR OF INFORMATION SCIENCE

CIT 3402: BIG DATA ANALYTICS

DATE: APRIL 2024 **TIME: 2 HOURS**

INSTRUCTIONS: Answer Question ONE and any other TWO questions.

QUESTION ONE (30 MARKS)

(a) Define the term Big Data and discuss any five of its characteristics. [6 Marks]

(b) Differentiate between data reporting and data analyzing. [2 Marks]

(c) Discuss the following terms as used in Big Data analytics.

i. Distribution
ii. Commodity hardware
iii. Distributed file system.

[6 Marks]

(d) Briefly explain how MongoDB handles any three of distributed systems challenges. [6 Marks]

(e) Briefly explain how MapReduce can process bank customers monthly transactions spanning across all branches in Kenya. [6 Marks]

(f) Outline any four differences between RDBMS and NoSQL [4 marks]

QUESTION TWO (20 MARKS)

(a) List any two data visualization tools for ordinal data. [2 Marks]

(b) Differentiate between the following types of data and give example in each case.

i. Nominal and ordinal data

ii. Interval and ratio data [8 Marks]

(c) Briefly explain the basic architecture of Hadoop platform [6 Marks]

(d) Discuss any two reasons why nonprofit making learning institution should invest on Big Data analytics tools. [4 Marks]

QUESTION THREE (20 MARKS)

a) Discuss the general process of Big Data analytics. [6 Marks]

b) Differentiate between vertical and horizontal scaling of storage system [2 Marks]

c) Briefly discuss any three filtering relational operators of Pig Latin tool. [6 Marks]

d) Briefly discuss any three of the following machine learning algorithms as used in Big Data analytics

i. Decision Tree

ii. Naive Bayes

iii. K-Means

iv. Random Forest

e) Dimensionality Reduction Algorithms [6 Marks]

QUESTION FOUR (20 MARKS)

(a) Briefly differentiate between HIVE and PIG [4 Marks]

(b) Briefly explain how clustering is used in statistical data analysis and state any three common use cases.

[6 Marks]

(c) Briefly explain how Collaborative filtering can be used for online customers recommendation systems.

[4 marks]

(d) Briefly explain security challenge of a distributed databases. Briefly outline how they are addressed in Cassandra or similar systems [6 Marks]

QUESTION FIVE (20 MARKS)

a) Discuss the concepts of intelligent data analysis (IDA) and state any four application areas.

[5 Marks]

b) Discuss any three of the following clustering techniques.

i. Partitioning Clustering

ii. Density-Based Clustering

iii. Distribution Model-Based Clustering

iv. Hierarchical Clustering

v. Fuzzy Clustering [6 Marks]

c) Using a diagram explain architecture of HDFS.state any four of its components

[9 Marks]