



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: info@must.ac.ke Email: info@must.ac.ke

University Examinations 2024/2025

FIRST YEAR FIRST SEMESTER FOR THE DEGREE OF MASTER OF SCIENCE IN DATA
SCIENCE

CCD 7103: ADVANCED DATA SYSTEMS

DATE: JANUARY 2025

TIME: 3 HOURS

INSTRUCTIONS: *Answer any three questions*

QUESTION ONE (20 MARKS)

- Explain the components of the Hadoop ecosystem, focusing on HDFS and Map Reduce. How do they contribute to distributed data storage and processing? (4 marks)
- Walk through the process of creating and optimizing a query execution plan for a SELECT statement that involves joining multiple tables. (6 marks)
- Describe the ACID properties in the context of data consistency and recovery in relational databases. How do NoSQL systems handle consistency differently? (4 marks)
- Use an example to elaborate the three V's (Volume, Velocity, and Variety) associated with Big Data. Describe how NoSQL fit into the Big Data landscape? (6 marks)

QUESTION TWO (20 MARKS)

- Design a schema for a data warehouse using either a Star or Snowflake schema for a retail company tracking sales, inventory, and customer interactions. Justify your choice of schema. (6 marks)
-

- b) Compare and contrast document-based and key-value NoSQL databases. What are their key features, and when might you choose one over the other? (4 marks)
- c) Describe the architecture of block chain technology and its relevance to data systems. How does a distributed ledger ensure data integrity and security? (4 marks)
- d) An e-commerce company experiences rapid growth in customer transactions, which causes performance issues with its existing relational database system. The company is considering moving from its relational database to a NoSQL system like MongoDB or Cassandra. How would you advise the company to approach this migration, and what factors should be considered when selecting the most appropriate NoSQL database? (6 marks)

QUESTION THREE (20 MARKS)

- a) A large retail chain wants to implement a data warehouse to store and analyze sales data from multiple store locations. The goal is to perform advanced reporting and OLAP-based analysis. Design a data warehouse schema for the retail chain using either a Star or Snowflake schema. Justify your design and explain how OLAP cubes could be used to generate sales reports. (5 marks)
- b) Explain how big data analytics can benefit your retail company by using the collected data. Provide specific examples of how analytics can enhance customer insights, inventory management, and marketing strategies. (5 marks)
- c) The company is keen on employing machine learning models to personalize recommendations for customers, optimize supply chain operations, and detect fraudulent activities. Describe how you would approach building and deploying machine learning models in a big data environment. Elaborate on challenges that may arise, and how would you address them? (5 marks)
- d) A logistics company is exploring the use of block chain to improve traceability and transparency in its supply chain operations. Design a block chain-based data storage system for this company, explaining how distributed ledger technology can ensure secure and tamper-proof tracking of products across the supply chain. How would smart contracts enhance the functionality? (5 marks)

QUESTION FOUR (20 MARKS)

You are a database administrator for a rapidly growing e-commerce company. The company's online store has experienced a surge in user activity, resulting in increased data volumes and performance issues with the current relational database system. You are considering the adoption of a NoSQL database solution to address these challenges.

- a) Explain why a NoSQL database might be a suitable choice for handling the increased data volumes and performance demands of your e-commerce platform. Provide specific examples of how different types of NoSQL databases (e.g., document-based, key-value, or column-family) can benefit your use case. (5 marks)
- b) Outline the key steps involved in the migration from a relational database system to a NoSQL database system. Discuss the challenges you might encounter during the migration process and how you plan to overcome them. (5 marks)
- c) Describe the considerations for data modeling and schema design in a NoSQL database compared to a relational database. Elaborate on how does schema flexibility in NoSQL databases impact your approach to managing data for product listings, user profiles, and order history? (5 marks)
- d) Security and data privacy are paramount in your e-commerce platform. Explain how you would ensure data security and access control in a NoSQL database environment. Elaborate the role of authentication, authorization, and encryption in safeguarding sensitive customer data. (5 marks)

QUESTION FIVE (20 MARKS)

- a) Provide examples of data transformation tasks that may be performed during ETL processes for data warehousing and data lake environments. (4 marks)
- b) Compare edge computing with cloud computing in terms of data management and real-time processing for Internet of Things (IoT) systems. What are the benefits of edge computing in latency-sensitive applications? (4 marks)

- c) An e-commerce company experiences rapid growth in customer transactions, which causes performance issues with its existing relational database system. The company is considering moving from its relational database to a NoSQL system like MongoDB or Cassandra. How would you advise the company to approach this migration, and what factors should be considered when selecting the most appropriate NoSQL database? (6 marks)
- d) Explain the importance of query optimization in improving the performance of the e-commerce website. Provide specific examples of how slow queries can impact user satisfaction and sales. (6 marks)