Question 1

a. You are a data scientist working for a leading healthcare organization. The organization is committed to improving patient care and medical research by harnessing the power of data analytics and machine learning. As part of your role, you have been assigned the responsibility to design a data management system to support the organization's data-driven initiatives.

Question 2

a. Imagine you are designing a database system for a university. The system needs to keep track of students, courses, and their enrolment status. Students can enrol in multiple courses, and each course can have multiple students enrolled. Additionally, you need to record student grades for each enrolled course. Professors teach these courses, and they may teach multiple courses. Courses are offered in different departments within the university.

Given the scenario of designing a database for a university, please explain how you would use Entity-Relationship modelling to represent the entities, attributes, and relationships in the system. Specifically, address the following:

a. What are the main entities that you would identify in this scenario, and why are they important for the database design?

b. How would you differentiate between attributes and entities in this context, and what attributes might you associate with each entity?

c. Describe the relationships between these entities in terms of E-R modelling. How would you represent the relationships between students, courses, professors, and departments?

d. Considering the enrolment status and student grades, how would you utilize cardinalities to define the relationships in the E-R diagram?

e. Could you provide suggestions for naming these entities, attributes, and relationships to ensure clarity and consistency in the E-R diagram?
Please provide a detailed explanation of your E-R diagram design based on this university database scenario, considering the various components of E-R modeling.

b. A company wants to create an online bookstore to manage book inventory, customer information, and orders. The database should store information about books, customers, authors, and orders. Each book can be written by one or more authors, and each author can have multiple books. Additionally, each order can contain

multiple books, and each book can be a part of multiple orders. Customers can place multiple orders, but each order belongs to a single customer. How can you design a relational database schema to fulfil these requirements?

Question 3

a. In a School of Computer Science teaching is based on units. Examples of unit of teaching include Operating Systems, Relational Model and Computer Architecture. A unit such as Operating Systems may be composed of smaller units (e.g Computer Architecture and Data Structures) and may be part of many larger units, such as Distributed Systems or Web Technologies. A unit is identified by its unit number (unitNo) and has a title as attribute.

Generate, with justification, a corresponding E-R diagram to represent each case.
Indicate clearly the attributes, the cardinality and the type of participation.

Derive relations from the diagram and express them as schemas.

b. You are tasked with processing a massive dataset containing social media posts and analysing the frequency of hashtags mentioned in these posts. Explain how you would use the MapReduce framework to accomplish this task. Provide a step-by-step description of the Map and Reduce phases, as well as any potential challenges or optimizations you would consider during the process.

Question 4

a. An e-commerce company wants to enhance its supply chain management and delivery processes by harnessing big data analytics from multiple sources, such as real-time inventory data, weather conditions, and customer demand patterns. How can they effectively utilize big data analytics to optimize inventory levels, minimize delivery times, and ultimately improve their overall operational efficiency?"

b. A financial institution wants to detect fraudulent activities and prevent financial crimes by analysing large volumes of transactional data, user behaviour patterns, and external risk factors. How can they utilize big data analytics methodologies, such as anomaly detection or network analysis, and NoSQL databases to process and analyse the diverse data sources in real-time?