Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
Cawangan Perlis, Kampus Arau
Case Study
DSC650 - Data Technology and Future Emergence
Mar 2025 – Jul 2025 (20252)

Case Study Title: "The Evolving Data Landscape of 'OmniHealth Analytics'"

Scenario:

OmniHealth Analytics (OHA) is a rapidly growing startup specializing in healthcare data solutions. Their primary goal is to leverage advanced data technologies to improve patient outcomes, optimize hospital operations, and accelerate medical research.

Initially, OHA started with a traditional relational database (RDBMS) to manage patient records and billing information. As their client base expanded and they began incorporating diverse data sources – including medical imaging (X-rays, CT scans, MRIs), wearable device data (heart rate, activity levels), genomic sequences, and unstructured clinical notes – their existing RDBMS infrastructure began to show significant limitations.

OHA now faces critical decisions regarding its data technology stack to support its ambitious future plans, which include:

- Developing a real-time patient monitoring system.
- Implementing predictive analytics for disease outbreaks.
- Creating a personalized treatment recommendation engine.
- Automating the analysis of vast amounts of medical literature.
- Ensuring high availability and fault tolerance for critical patient data.
- Providing intuitive data discovery tools for medical researchers.

You have been hired as a data technology consultant for OmniHealth Analytics. Your task is to analyze their current challenges and propose a comprehensive data technology strategy, drawing upon the concepts discussed in DSC650.

Assignment Questions:

Please answer the following questions, providing detailed explanations and justifications based on the lecture material.

1. Current Challenges & RDBMS Limitations:

    a. Identify at least three specific limitations of OHA's initial RDBMS infrastructure that would become apparent when dealing with the new, diverse, and large-scale healthcare data. Explain why each limitation is problematic in this context.

    b. Discuss how the characteristics of "Big Data" (Volume, Velocity, Variety, Veracity, Value) apply to OmniHealth Analytics' evolving data landscape. Provide concrete examples for each characteristic.

2. Proposed Core Data Infrastructure:

    a. Recommend a core distributed file system for OHA's large-scale, diverse data. Justify your choice by explaining its key features and how they address OHA's needs (e.g., scalability, fault tolerance, handling unstructured data).

    b. Explain the architectural components of your recommended distributed file system (e.g., NameNode/DataNode in HDFS). Describe how data is stored and managed within this system.

Submission Guidelines:

- Your answers should be well-structured, clear, and concise.
- Ensure your responses directly reference concepts and terminology from the lecture material.
- Provide justifications for all your recommendations.
- The total length should be appropriate for a comprehensive analysis (e.g., 1500-2500 words, excluding references).
- Cite any external sources if used, in addition to the provided lecture material.

Evaluation Criteria:

- Understanding of Concepts: Demonstrated grasp of data technology concepts from the lecture.
- Application to Case Study: Ability to apply theoretical knowledge to a practical scenario.
- Justification & Reasoning: Sound arguments and logical explanations for recommendations.
- Clarity & Structure: Well-organized, easy-to-read, and professional presentation.
- Completeness: Addressing all parts of each question thoroughly.