Assignment 1

Voningo Flora Makhubele Pecego

ST10483982

School of Computer Science, Varsity College

INSY6112: Information Systems 1B

Lecturer: Matthew Prinsloo

Group 3

Date: 26 August 2025

Table of Contents

Question 1

The scenario provided outlines a social media platform that is responsible for collecting and managing large amounts of real-time data, such as text posts, videos, likes, and shares, to support a massive user base (The IIE, 2025). As a result, the system faces several key challenges, including exponential data growth, complex data structures, the need for real-time analytics, and maintaining a seamless user experience (The IIE, 2025). Based on the nature of the data required to support and maintain the platform, the following report recommends the adoption of a NoSQL database.

1. Definition of the Recommended Database Type

NoSQL (Not Only SQL) refers to a non-relational database system that uses modelling techniques such as key-value stores, graph databases, and document data models (Mohamed et al., 2014) to store unstructured or semi-structured data. As such, NoSQL databases are designed from the ground up to handle large-scale, high-performance, non-relational data stores (The IIE, 2025). Furthermore, NoSQL databases offer scalability, flexible schema design, high availability, and the ability to process real-time analytics (Oracle, 2021).

2. Motivation for Recommendation

The following table lists three benefits of NoSQL databases that motivate why it is most suitable for the social media platform (Oracle, 2021):

Benefit	**Description**
Scalability	NoSQL databases support horizontal scaling across multiple servers using a process called “sharding.” This horizontal scaling, as opposed to the vertical scaling offered by relational databases, requires less additional processing power and memory and has the ability to efficiently handle extremely large amounts of data, even as it grows exponentially. In this way, their scalability makes NoSQL databases suitable for real-time web applications that must manage large data volumes and high traffic from millions of users, while requiring low latency for seamless user interaction.
Flexibility	Unlike relational databases, which require predefined schemas and work best with structured data (The IIE, 2025), NoSQL databases can store structured, semi-structured, and unstructured data such as text, images, and videos, making them suitable for a social media platform. Furthermore, NoSQL databases offer flexibility with regard to defining schemas. As a result, data models can be defined as the application evolves. For example, new content types or features can be added without redesigning the database schema. This allows the social media application to innovate, improve, and develop more rapidly.
High Performance	NoSQL databases are equipped to handle high-velocity, high-volume, and high-variety data (The IIE, 2025), all of which are characteristics of a social media platform. Such a platform must handle millions of users generating large

	amounts of data, using various data structures, while also providing real-time analytics and feedback to ensure a seamless user experience (The IIE, 2025). This is possible because NoSQL can handle high traffic while still offering low latency or fast response times. For example, it prevents delays in user engagement metrics such as likes, comments, and shares appearing on the platform, while also providing instant notifications.

3. Discussion of the Kinds of Data Stored

The database would be required to store data about the user, for example, usernames, email addresses, and full names. Additionally, it would store unstructured data that is shared or posted on the platform, such as text posts, videos, and images. It would also store user engagement data such as likes, reactions, and comments, as well as data regarding user relationships, such as a user's following or connections. Lastly, data about the user's interaction with the platform itself would be stored, such as login times and activity logs.

4. Types of NoSQL Databases

The following table lists and defines four types of NoSQL Databases:

NoSQL Database Type	**Definition**
Document Store	Document store databases use the document model format, such as JSON (JavaScript Object Notation) (The IIE, 2025), to store data as flexible, semi-structured objects that are hierarchical in nature. This enables them to evolve with the needs of an application (AWS, n.d.). They are well suited to storing user profiles and content such as posts and comments, that must evolve continuously with new platform features. An example of a document store database is MongoDB (AWS, n.d.).
Key-Value Store	Key-value store databases, such as Amazon DynamoDB (AWS, n.d.), store data as a collection of key-value pairs (The IIE, 2025). They make use of keys, which act as unique identifiers, and values of varying complexity. They also offer horizontal scaling at a higher level than other types of NoSQL databases (AWS, n.d.). Key-value stores are commonly used for features such as user sessions, caching, and notifications, where consistent performance and low latency are critical (AWS, n.d.).
Graph Store	Graph databases, such as Amazon Neptune (AWS, n.d.), use nodes to store entities and edges to represent and store relationships between them (The IIE, 2025). This makes them well suited for building and running applications that work with highly connected datasets, such as social networking platforms and recommendation engines. For example, the "people you may know" feature on Facebook uses a graph-based recommendation engine to suggest profiles to users, thereby enhancing personalization.

In-Memory Store	In-memory databases aim to remove the need to access disks or SSDs (AWS, n.d.), as data is stored directly in system memory (RAM). They are well suited for applications that require microsecond response times and that must handle large traffic spikes, supporting features such as real-time analytics instant feed updates. For example, Amazon MemoryDB for Redis is an in-memory database that offers very low read/write latency.

5. The Three Vs of Big Data in the Scenario

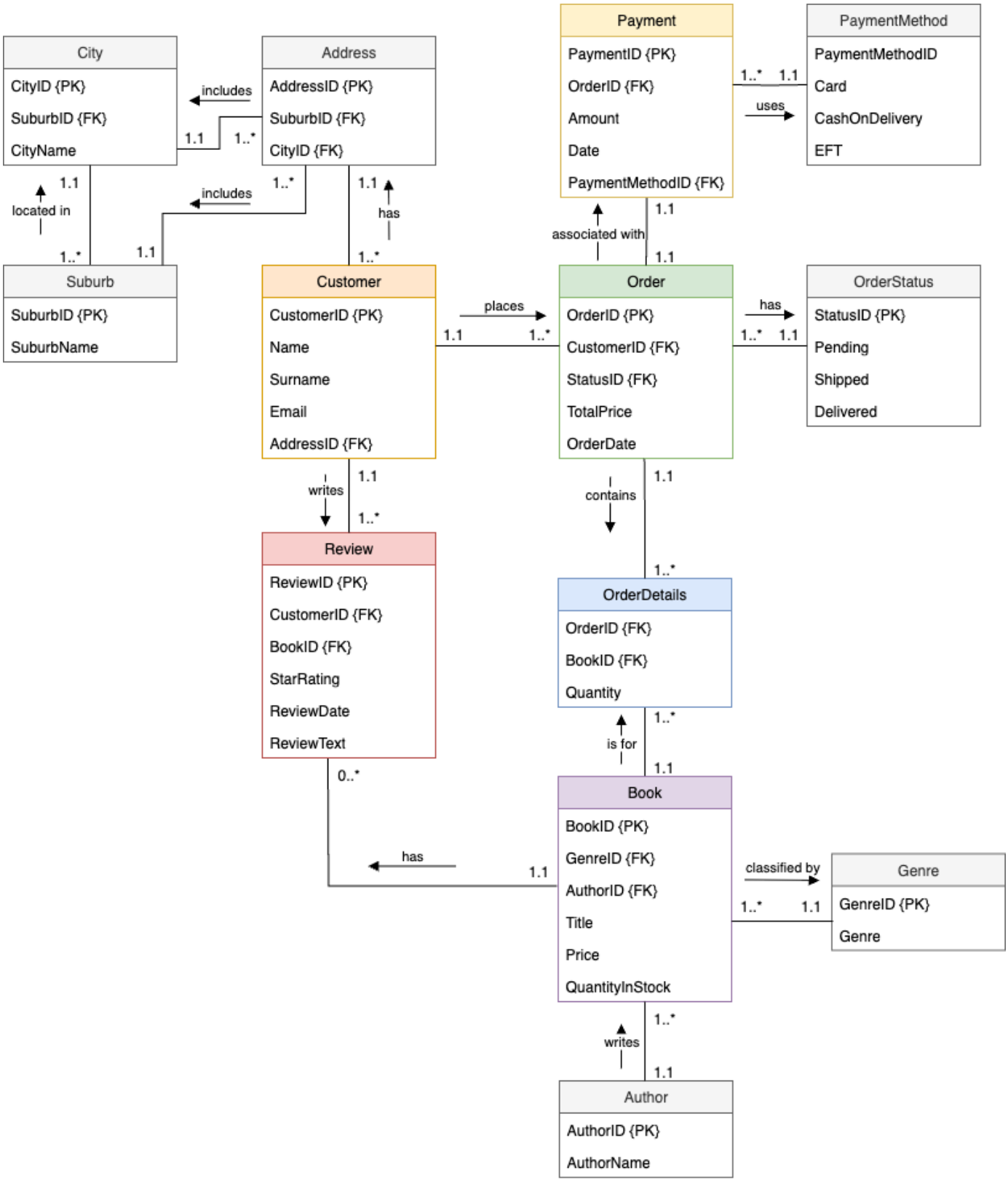
1. **Volume** - Millions of users interact on the social media platform, resulting in millions of posts, comments, likes, and videos created daily. This leads to a massive volume of data that continues to grow exponentially, particularly as new users join the platform or new features are added (The IIE, 2025).

2. **Velocity** - When users engage with the platform, large amounts of data are generated automatically — for example, likes per second and live comments — and this data must often be processed almost instantly. Real-time processing ensures that analytics, notifications, and feeds are updated without delay, thereby maintaining low response times and delivering consistent performance (The IIE, 2025).

3. **Variety** – Social media platforms handle a wide variety of data structures, as users can share and interact with different types of content such as videos, images, text, and live streams. As a result, a NoSQL database is required to process this mix of structured, semi-structured, and unstructured data efficiently, while ensuring both scalability and real-time performance to maintain a seamless user experience (The IIE, 2025).

Question 2

Entity Relationship Diagram



References

AWS, [n.d.]. *What is NoSQL?*. [online] Available at: <https://aws.amazon.com/nosql/> [Accessed 25 August 2025].

Mohamed, M, Mohammed, I and Obay, I. 2014. Relational Vs. NoSQL databases: A survey. *International Journal of Computer and Information Technology*, [e-journal] 3. pp.598. Available at: <https://www.researchgate.net/publication/263272704_Relational_Vs_NoSQL_databases_A_survey> [Accessed 26 August 2025].

OpenAI. 2025. Chat-GPT (Version 1.2025.218). [Large language model]. Available at: <https://chatgpt.com/share/68adfb1e-0670-8000-a7c6-ae1397597222> [Accessed: 26 August 2025].

Oracle, 2021. *What is NoSQL?*. [online] Available at: <https://www.oracle.com/africa/database/nosql/what-is-nosql/> [Accessed 25 August 2025].

The IIE. 2025. Information Systems 1B [INSY6112 Module Manual]. The Independent Institute of Education: Unpublished.

Disclosure of AI Usage

Section(s) within the assessment in which generative AI was used
Assignment Question 1 and Assignment Question 2.

Name of AI tool(s) used
ChatGPT Version 1.2025.218

Chat Link
<https://chatgpt.com/share/68adfb1e-0670-8000-a7c6-ae1397597222>

Purpose/intention behind use
AI was used for the following to perform the following tasks

- Brainstorming ideas,
- Recommending sources,
- Identifying knowledge gaps,
- Drafting research plans and outlines,
- Correcting grammar and spelling,
- Evaluating answers, and
- Identifying errors.

Date(s) in which generative AI was used
AI was consulted from 24 August 2025 – 26 August 2025
