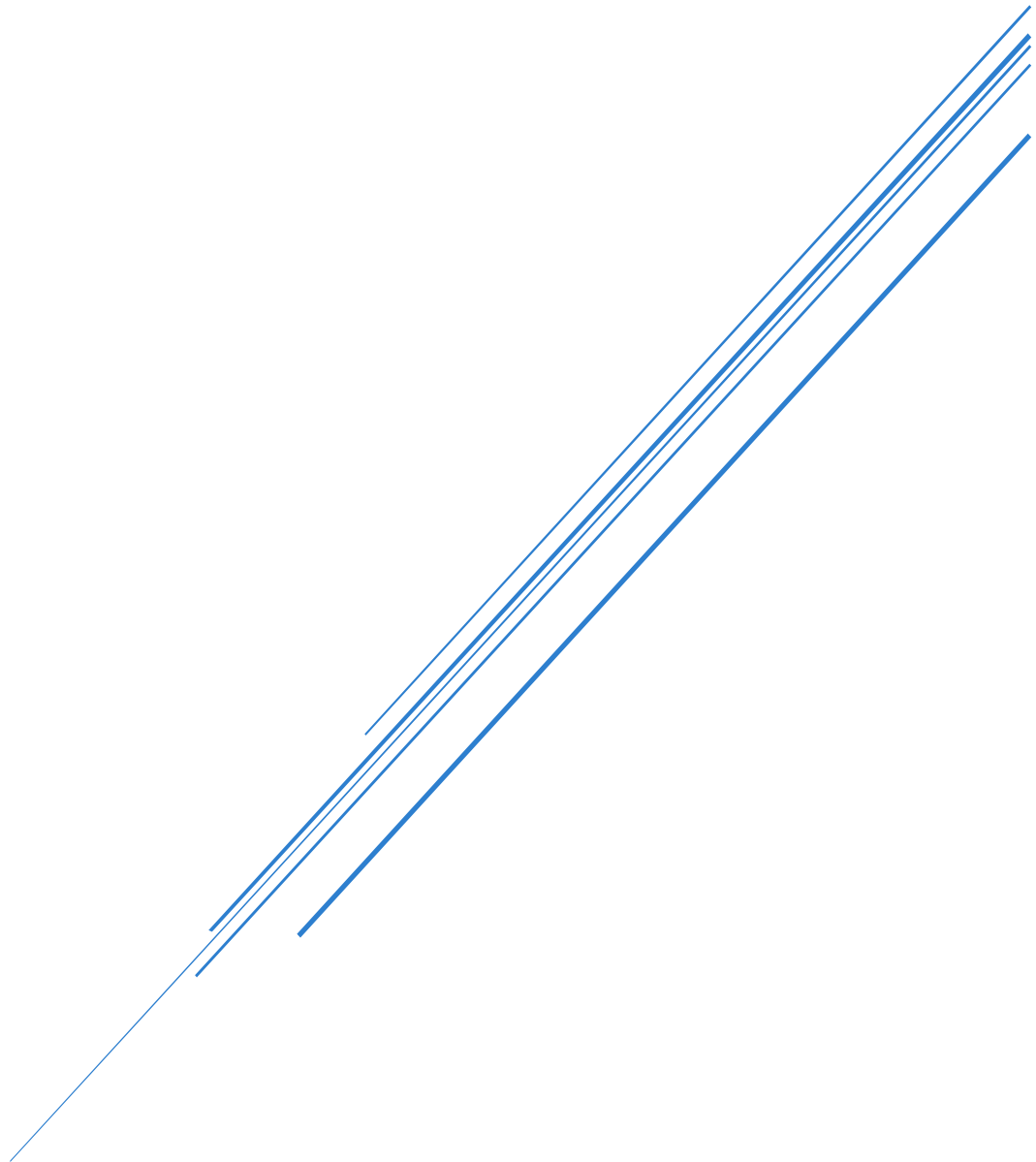# INFORMATION SYSTEMS 1B: ASSIGNMENT 1

Student Number:ST10467645
Full Name: Kandyce Jade Smit

Lecturer: Prof. Lucas Khoza
Module Code: INSY6112

Contents

Question 1

- I would recommend a NoSQL database for data storage. A NoSQL database is a non-relational database, and it makes data accessible and stores the data, using one of several flexible formats, like key-value pairs, document-based structures or graphs. Unlike relational databases, the NoSQL databases do not have to be designated with a fixed schema and can accommodate unstructured or semi-structured data efficiently. NoSQL databases were built for large-scale, dynamic applications needing performance and scalability to support their user base. (TechTerms, 2013).

- NoSQL is the best fit for this platform because of the nature of the data and the activities this system needs to support. These are the following reasons why:

  1.) **Handles different kinds of data easily**
  On social media platforms, users post various types of content; text, images, videos, GIFs, and of course also things such as live streams or VR experiences. A NoSQL database does not need to have a single structure to store everything which makes it an ideal storage method for this type of changing and agile data. There will be no need to redesign the schema every time there is a new post type or feature. (Refonte Learning, 2025).

  2.) **Designed to be able to handle and grow with lots of data**
  On social media, the database needs to be able to scale because millions of users are posting, commenting, liking and sharing continuously. NoSQL databases can scale horizontally, meaning new servers can be added easily to keep performance stable as usage grows. (Refonte Learning, 2025).

  3.) **Real-time performance and speed**
  When social media needs things to turn around quickly, especially for things like likes, comments, or trending posts, NoSQL stores and processes data fast, so users see updates in real time. This keeps the platform smooth and the user experience consistent. (Refonte Learning, 2025).

- For the social media platform scenario, the platform will need to store different types of data to support features like user profiles, content, interactions, and analytics. Here's how the data would be categorised:

  1.) **User Profiles (Structured/Semi-Structured Data)**
  User Profiles would include usernames, emails, user preferences, follower/following lists, user defined settings, and geolocation, and all these fields will be classified as semi-structured and a document structure format like JSON will be used to store. (PMC, 2020; MongoDB, n.d.).

  2.) **Content Posts (Unstructured/Semi-Structured)**
  Content posts could include text posts, photos, videos, GIFs, presentations, live streaming, or virtual reality content, and they would each have their own metadata associated. Each one of the posts does not have a fixed structure so document storage would be the best solution for storage. (PMC, 2020; MongoDB, n.d.).

  3.) **Interactions and Engagements (Structured/Event Data)**
  Interactions would include likes, comments, shares, view counts, timestamps, and every interaction would include metadata. Each one of the interactions would

generally have a high volume, be updated frequently, and be event/time series data. (PMC, 2020; MongoDB, n.d.).

4.) Trend and Analytics Info (Aggregated Metrics)
This will include real-time metrics likes per minute, trending hashtags statistics, engagement rate indicators and use in reporting dashboards, recommendation engines and insights. (PMC, 2020; MongoDB, n.d.)

- The four types of NoSQL Databases:

1.) Document Databases
With document databases, the data is stored in JSON-like documents. Each of the records in a document database is a document, and a document allows each one of these records to have its own structure and metadata. For example, in a social media platform, a user profile or a post (which has text, an image link, tags) could be created in a single document. Whenever needed, we could retrieve the whole document (e.g. relation) at once in a single read. When we use document databases, we have the flexibility of building documents without modifying tables as new features are introduced. (AWS, n.d.).

2.) Key-Value Databases
Key value databases usually represent a unique key matching one value payload. Key value databases are very quick since we can look up the value exactly by key which is useful for something simple, e.g., to store session tokens or user activity caches or notification settings. Key value databases are simple and quick, so they are useful for some very high load functions. For example, to retrieve a user's preferences quickly, or to use caching for a user's feed. (Google Cloud, n.d.).

3.) Column-oriented databases
Data in column-oriented databases can be stored in flexible tables and uses columns instead of rows. Storing the data in rows always means each row must have the same structure and columns, using columns means each row has completely different columns and scale across clusters using horizontal sharding. Column-oriented databases are useful for processing really big data sets, e.g., analytics logs or timelines of users' feed; how many metrics does each user have, and through time; we could write, and read values tens of metrics quickly even when working with a very small subset of metrics, the column subset could be enormous but it's just a partial load of the whole data set. (Google Cloud, n.d.).

4.) Graph Database
Graph databases model data as nodes and edges representing the utilities and their relationships. In a social network for example users, posts, likes and shares are all relationships in a graph. This model for exploration allows querying to return "friends of friends" recommendations, and trending content networks efficiently. Graph databases optimize traversing complex relationships in real-time (e.g., calculating mutual connections). (Medium, 2023).

- The three Vs of big data include:

1.) Volume
Volume refers to the massive amount of data that big data systems are designed to handle, which perfectly matches this platform's needs (The Knowledge Academy, 2025). The magnitude of data related to this social media platform is extraordinarily large because it has millions of users that are actively generating and interacting

with posts, images, videos, likes, shares, and comments every minute of every day. The volume of data will climb exponentially as users continue to grow and as more features such as VR posts are introduced. However, since the volume of data is so immense, the system needs to take advantage of horizontal scalability to deal with a big data structure.

2.) Velocity

Velocity is the rate of data creation, processing and analytics (The Knowledge Academy, 2025). In this situation, velocity is an important consideration because of the low latency requirement for proficient user satisfaction. All the actions linked to user base postings, liking, and commenting occur continuously and must be engaged so they can be acted upon quickly without delay. The rate of analysis on the trends (hashtag or post popularity) of a single social media post must be updated instantaneously.

3.) Variety

Variety is referring to the variety of different kinds of data that may be embedded in big data systems (The Knowledge Academy, 2025). The platform consists of a variety of data formats and structures. There are structured formats such as user profiles and unstructured formats like videos, gifs and live streams. The storage structure cannot expect to only work with structured data, and a very defined/rigid schema for unstructured data. When looking at text, multimedia and analysis of user interaction in social media, we are talking about variety in an incredibly complex environment.

Question 2

Entity Relationship Diagram (ERD) using Unified Modelling Language (UML) notation according to the background information provided in the assignment:
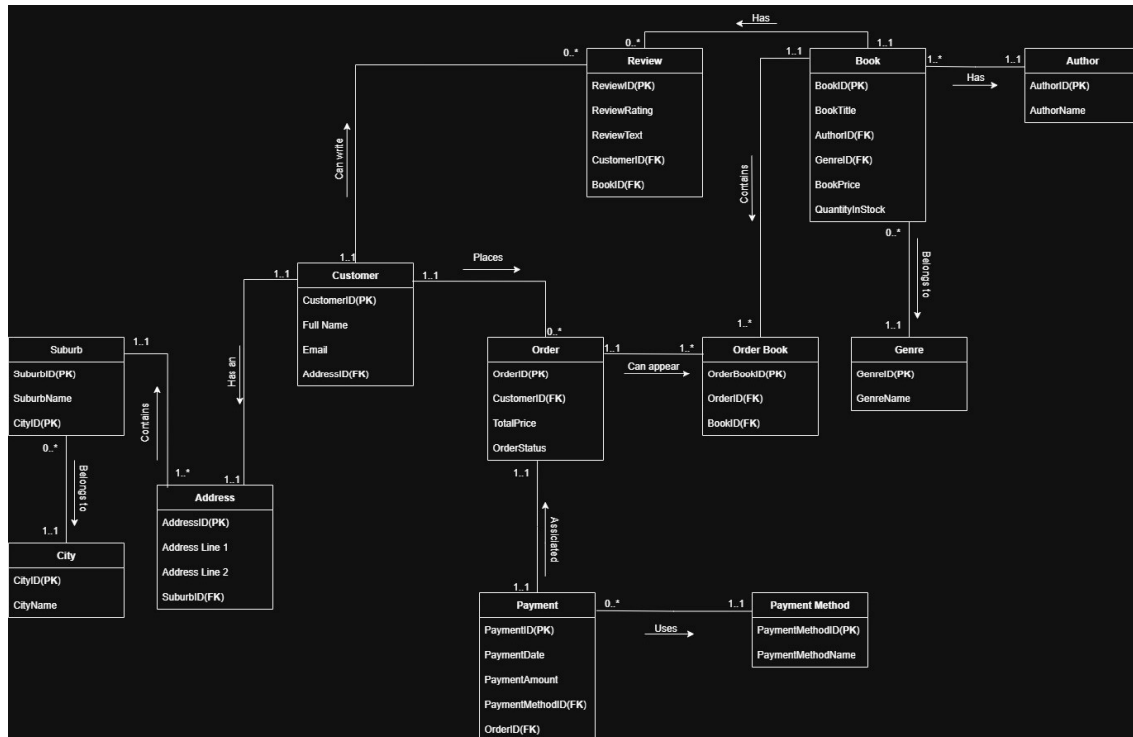
The entity relationship diagram was created using draw.io (Draw.io, n.d.).

Bibliography

AWS, n.d. *What is a document database?*. [Online]
Available at: https://aws.amazon.com/nosql/document/
[Accessed 05 August 2025].

Draw.io, n.d. *Draw.io online diagram editor*. [Online]
Available at: https://app.diagrams.net/?src=about
[Accessed 19 August 2025].

Google Cloud, n.d. *What is a NoSQL database?*. [Online]
Available at: https://cloud.google.com/discover/what-is-nosql
[Accessed 05 August 2025].

Medium, 2023. *How to choose a database for a social media app*. [Online]
Available at: https://medium.com/%40nebulagraph/how-to-choose-a-database-for-a-social-media-app-e8072a950c9c
[Accessed 05 August 2025].

MongoDB, n.d. *What is NoSQL?*. [Online]
Available at: https://www.mongodb.com/resources/basics/databases/nosql-explained
[Accessed 05 August 2025].

PMC, 2020. *Big data analytics meets social media: A systematic review of techniques, open issues, and future directions*. [Online]
Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC7553883/
[Accessed 05 August 2025].

Refonte Learning, 2025. *SQL vs NoSQL: Choosing the right database for your project*. [Online]
Available at: https://www.refontelearning.com/blog/sqlvs-nosql-choosing-the-right-database-for-your-project
[Accessed 04 August 2025].

TechTerms, 2013. *NoSQL definition*. [Online]
Available at: https://techterms.com/definition/nosql
[Accessed 04 August 2025].

The Knowledge Academy, 2025. *3 Vs of big data with example*. [Online]
Available at: https://www.theknowledgeacademy.com/blog/big-data-3v/
[Accessed 06 August 2025].