Student Number: ST10460086

Name: Reece Mikhael Moodley

Module Name: INFORMATION SYSTEMS 1B

Module Code: INSY6112

Assessment Type: ASSIGNMENT 1

**Question 1:**

**Introduction:**

This report will discuss the types of databases and key challenges large scale platforms face when handling big data. From the scenario provided, a **NoSQL Database** would be the most suitable to address the challenges of the social media platform. According to (Lutkevich, 2021) a database stores various forms of information that can be processed, analysed and managed in a system.

**What is a NoSQL Database?**

A NoSQL database is a storage system designed to store and manage massive amounts of data in non-tabular formats. A NoSQL database can accept structured data, its strength is in managing a wide range of data types mainly semi-structured and unstructured data in flexible, schemeless formats. This is a non-relational data store (Linda Nguyen, 2024).

**Reasons for Recommendation:**

* **Scalability** – The platform *requires* a system that can scale *horizontally*. The relational database can only scale vertically. This is to accommodate growth of the platform when it requires more servers. Vertical scaling only provides upgrades of resources to a single server.
* **Data Volumes** – One of the key challenges noted in the scenario is a high volume of data. The NoSQL database was designed to handle large volumes of data. A relational database struggles with vast amounts of data.
* **High Performance** – The social media platform requires real-time analytics and instant updates of new content and data. NoSQL databases often prioritize high performance and low latency over consistency and reliability. This characteristic is more suited to the system’s needs.
* **Flexible Schemas** – Due to the nature of a social media platform, most data will be unstructured. Photo’s, files and videos have no specific structure; hence it may not store neatly in tables. Flexible data types, such as JSON, XML, or binary data are supported by NoSQL databases. A relational database only supports structured data.

**Expected forms of data that will be stored in the database:**

**User Data** – A profile of the user, usually their username, full name, email address, passwords and bios in the form string data types.

**Content Data** – Content generated by the users, usually text, gifs, videos and images. These file types can vary greatly between each other. A regular structured table may not effectively analyse this type of data.

**Interaction Data** – On a user’s posted content, other users can like, comment, share and react. This data will represent the engagement the content has generated.

**Analytical Data** – Traffic metrics is an essential form of data content creators and businesses will greatly benefit from. This can include page views, new visitors, and returning visitors. Most of this information will be numerical values.

**Types of NoSQL Databases:**

* **Document Store** – According to (Linda Nguyen, 2024) documents such as JavaScipt Object Notation and BSON, manage, store and retrieve data in document databases. Nested fields, arrays and other structures organise data within each document. Documents have no predefined schema, which allows for greater flexibility in how data is expressed.
* **Key-value Store** – Key-value Databases hold data in the form of key-value pairs. Each key is a unique identifier such as a name or ID that relates to a value, which can range from a simple string to a more complicated data structure, such as a list or an object. Key-value databases are designed for rapid, efficient, and simple key lookups. Complex query operations are limited.
* **Column Store** – Data is stored column by column instead of rows. Column families are used to organise related information. Different number of columns can be found in each row. This flexibility enables column-oriented databases to accommodate enormous volumes of data with diverse schemas.
* **Graph Store** – Graphs that consist of nodes and edges are used to store, manage and query data. Entities are represented by nodes or vertices. Node relationships are represented by Edges. Starting at one node, a query moves across a chain of relationships to reach other connected nodes.

**The Three Vs of Big Data:**

**Volume** – According to (AltexSoft, 2021) this refers to vast amounts of data collected from a variety of sources, including IoT devices, social media, text files, business transactions, and more. The social media platform is required to handle huge amounts of content such as text, images, videos and gifs. Each file may not require excessive amounts of bytes to store them, however there will be millions of users constantly updating and adding new content that will drastically increase the range of information on the system.

**Velocity** – According to (AltexSoft, 2021) users demand real-time processing speeds. Velocity refers to the rate of speed at which data can be generated and processed. This is highly beneficial for the system as any delays can result in poor engagement for the company. Real-time analytics requires automatic generation for faster creation of data.

**Variety** – Illustrates the diversity of Big Data. Most big data is unstructured. In a social media post, the system can expect content such as text, images, videos and gifs. The user-experience aspect of the system handles various interactions that require an unstructured database to handle these requests.**Question 2:**

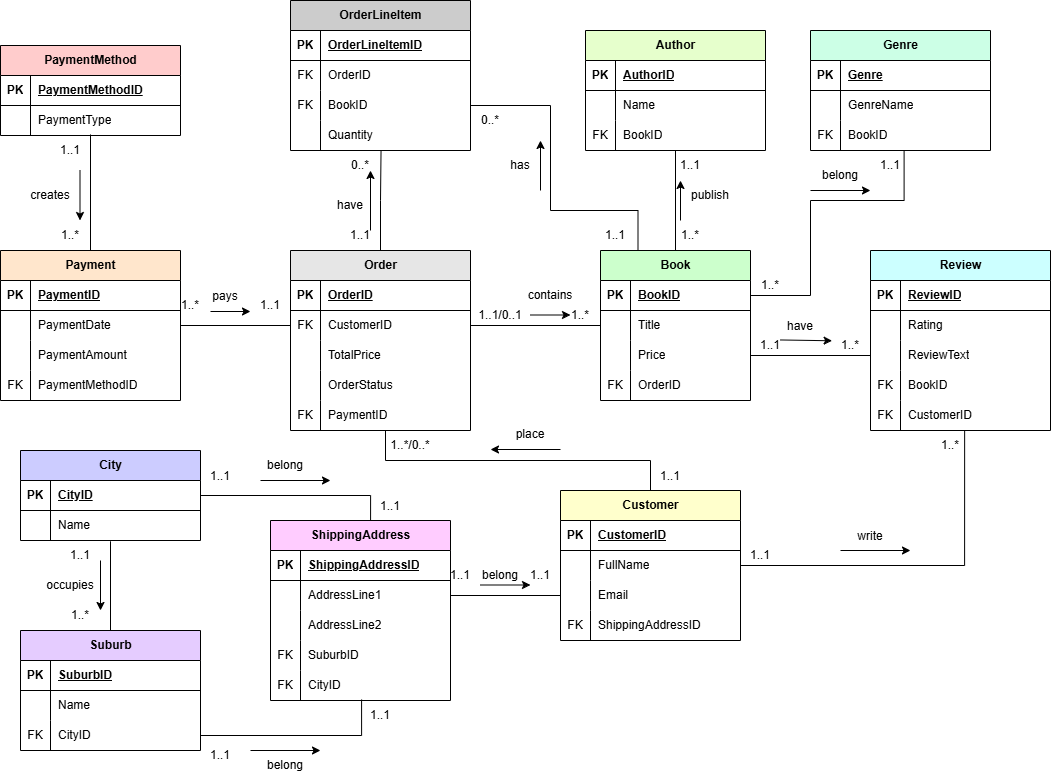


Figure 1: ERD

**Reference List:**

AltexSoft (2021). *Big Data Analytics: How It Works, Tools, and Real-Life Applications*. [online] AltexSoft. Available at: <https://www.altexsoft.com/blog/big-data-analytics-explained/> [Accessed 22 Mar. 2025].

Linda Nguyen (2024). *NoSQL Databases Visually Explained with Examples*. [online] AltexSoft. Available at: <https://www.altexsoft.com/blog/nosql-databases/> [Accessed 22 Mar. 2025].

Lutkevich, B. (2021). *What is a Database? Definition from SearchDataManagement*. [online] SearchDataManagement. Available at: <https://www.techtarget.com/searchdatamanagement/definition/database> [Accessed 21 Mar. 2025].

The IIE. 2012. Information Systems 1B [INSY6112/w Module Manual]. The Independent Institute of Education: Unpublished.