INSY6112 Assignment 1

Student number: St10493814

Contents

[Question 1: 2](#_Toc207101029)

[Question 2: 5](#_Toc207101030)

[Bibliography 6](#_Toc207101031)

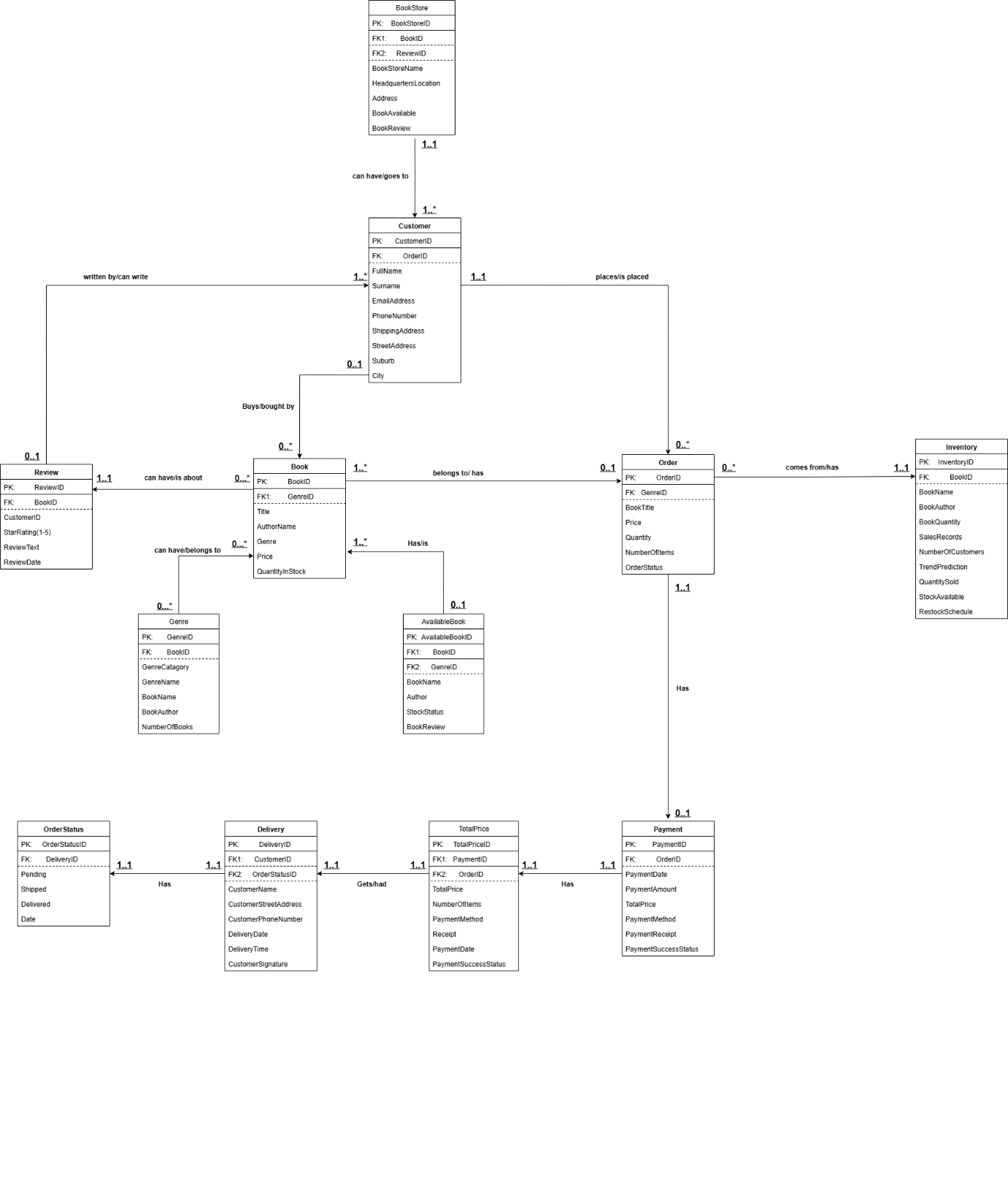
# Question 1:

1. The NoSQL (Not Only SQL) database is a database that is different from the traditional relational model. It is used for massive amounts of unstructured and semi-structured data and allows for seamless organization data and scalability. NoSQL databases do not have a universal query language meaning each NoSQL database approaches query languages uniquely. (Keita, 2022)
2. The NoSQL database is the recommended database because of its horizontal scalability, dynamic schema, Distributed and high availability. It has multiple benefits in relation to social media usage and functionality. Horizontal scalability refers to a systems capacity and ability to expand by adding additional nodes/servers instead of upgrading the current servers. (Custer, 2023) It can be cost effective and if the system is configured perfectly then there will be significantly less downtime, which is required for social media performance. Dynamic schema’s flexibility allows new fields to be added without disrupting existing fields of data, allowing seamless performance. Additionally, it is availably distributed for many users, its ability to handle large amounts of users, data while maintaining its performance is important so that applications used do not crash from the sheer amount of data being processed. These are also reinforced by recovery protocols and failover capabilities in order to recover from system issues faster, reducing the total downtime. (Keita, 2022)
3. This database can contain various types of data such as
   1. user-generated content
   2. multimedia data
   3. user profiles and their relationships
   4. Analytics data

User-generated content is content created by users on websites/applications that enable them to create various media that other users can interact with as well E.g. posts, comments and likes. Multimedia data is a collection of various types of data used to make content. Different medias store, deliver and use data differently E.g. pictures, videos, gifs and Instagram boomerangs (GG, 2025). User profiles and their relationships describe an individual’s information such as personal log in information and the type of content they engage with on the application (Carter, 2025). Their relationships can be between their followers, their following or the users content engagement. Lastly, analytics data refers to the analysis of unprocessed data and its results. In the context of social media, it is used to automate algorithms and trends pertaining to the user’s interests. These stats are used to improve data processing, management and user experience. (The Investopedia Team, 2025)

1. The main types of NoSQL databases are document databases, key-value databases, wide-column stores and graph databases. (Geeksforgeeks, 2025) The way Document-oriented database store information resembles JSON (JavaScript Object Notation) objects. This database has fast dynamic schemas and work primarily with unstructured and semi-structured data while also supporting nested structures that assist in displaying complex relationships and hierarchical information. Documents will often include pairs of values and fields which include strings, arrays, numbers, Boolean etc. MongoDB is an example of a document-oriented database. (Keita, 2022) Key-value databases contain unique keys and values. They are easier to understand and are used for session management and caching real-time metrics because it can efficiently read and write information stored in memory. It is highly scalable which is required for seamless social media performance. An example is Amazon DynamoDB. A wide-column store organises data in dynamic columns and tables rather than rows for increased retrieval efficiency. The column compression technique can be utilised to decrease storage usage for improved performance. Some examples are Apache Cassandra and HBase. (Keita, 2022)Graph databases store data as nodes(users) and edges(follows/relationship). Nodes store data about people, places and more while edges store data about the relations between the nodes. Complex relationships and their patterns are easily dissected and managed. And this is essential for social media as it enables for nodes and edges to be connected for increased socialisation between users, and the media available. An example is Amazon Neptune. (MongoDB, n.d.)
2. The three Vs of big data are Volume, Velocity, and Variety. *Volume* is how much data there is, velocity refers to how fast data is processed and variety refers to the various types of data (Lutkevich, 2023). Social media platforms contain large amounts of data that require advanced storage and processing programs. Things such as post, likes, comments, video/image uploads, shares and saves etc require real-time processing and so in turn social media platforms have enormous scaled out data management and storage facilities. A high Velocity is required for real-time processing and efficient management for updated like counts, notifications and comments and uploads so that users can engage seamlessly with the content. The application must be able to also handle sudden spikes of usage and interacts when there is an event so that it doesn’t crash. Finally, Variety refers to the types of data such as images, videos, livestream, comments, voice messages etc (Studocu Ai , 2025). Supporting a variety of data requires advance data management and the use of analytics to be able to decipher unstructured data. Advance systems are required to be able to process, manage and support all types of media.

# Question 2:

Entity Relational Diagram: 

# Bibliography

Carter, M., 2025. *Understanding User Profiles: An In-Depth Guide to 8 Key Dimensions.* [Online]   
Available at: https://boardmix.com/articles/what-is-user-profiles/  
[Accessed 24 August 2025].

Custer, C., 2023. *Vertical vs. horizontal scaling: What’s the difference and which is better?.* [Online]   
Available at: https://www.cockroachlabs.com/blog/vertical-scaling-vs-horizontal-scaling/  
[Accessed 11 August 2025].

Geeksforgeeks, 2025. *Introduction to NoSQL.* [Online]   
Available at: https://www.geeksforgeeks.org/dbms/introduction-to-nosql/  
[Accessed 11 August 2025].

GG, 2025. *Multimedia Database.* [Online]   
Available at: https://www.geeksforgeeks.org/dbms/multimedia-database/  
[Accessed 24 August 2025].

Keita, Z., 2022. *NoSQL Databases: What Every Data Scientist Needs to Know.* [Online]   
Available at: https://www.datacamp.com/blog/nosql-databases-what-every-data-scientist-needs-to-know  
[Accessed 11 August 2025].

Lutkevich, B., 2023. *3 V's (volume, velocity and variety).* [Online]   
Available at: https://www.techtarget.com/whatis/definition/3Vs  
[Accessed 13 August 2025].

MongoDB, n.d. *What is NoSQL?.* [Online]   
Available at: https://www.mongodb.com/resources/basics/databases/nosql-explained#:~:text=NoSQL%20databases%20come%20in%20a,data%20and%20high%20user%20loads.  
[Accessed 11 August 2025].

Studocu Ai , 2025. *Discuss how the three Vs of big data apply to this social media.* [Online]   
Available at: https://www.studocu.com/en-za/messages/question/11930464/discuss-how-the-three-vs-of-big-data-apply-to-this-social-media-platform-scenario  
[Accessed 13 August 2025].

Tang, Y., 2023. *What Is A Dynamic Schema?.* [Online]   
Available at: https://zilliz.com/blog/what-is-dynamic-schema  
[Accessed 11 August 2025].

The Investopedia Team, 2025. *Data Analytics: What It Is, How It's Used, and 4 Basic Techniques.* [Online]   
Available at: https://www.investopedia.com/terms/d/data-analytics.asp  
[Accessed 24 August 2025].