Information Systems 1B
Assignment Name: Databases
Module Code: INSY6112
Student Number: ST10466840
Student Name(s)Blessing Jafari
Surname: Mbalaka
Date Submitted 2025/08/25
Github link:
https://github.com/VCPTA/bca1-insy6112-practical-assignment-submission-ST10466840.git

---

Question 1:

Executive Summary
SQL (Structured Query Language) databases have been the industry standard for many years, and their structured benefits have been insurmountable. However, the age of big data has arguably required a more adaptive data storage approach. This report will be highlighting the necessitated adoption of Not only Structured Query Language (NoSQL) Databases for social media platforms.

Defining NoSQL
The recommended database type is No-SQL and this section will be highlighting these in conjunction to the key challenges being considered. According to Vettor et al., 2020, cited in IIE, 2025, the term NO-SQL is defined as a high-performance data store, which uses a resilient and scalable approach to storing data. This approach differs from the conventional joining of normalised data. Instead, No-SQL, uses unstructured and semi-structured data, which Vettor et al further elaborate upon as including, and not limited to being stored in key-value pairs JSON (Javascript Object Notation) documents. JSON (Javascript Object Notation) is useful and can be used with DMBS (Database Management Systems) such as MongoDB.

The Motivation for NoSQL

- The scenario above describes a social media platform that needs to have storage provisions for complex data structures, high volume of data, real-time analytics. Thomas (2023), argues that traditional SQL and its architectural design creates performance handicaps which inhibit its ability for real-time analytics; impediments

which are not seen in NoSQL databases. Therefore, NoSQL mitigates this constraint prevalent in SQL databases.

- The requirement for the database to be horizontally scalable is a characteristic of NoSQL databases (Exponent, 2023). In addition to this, these databases are easy to set up.
- There is a concern pertaining to how the data structures should not be constrained by rigid schema constraints. The solution to this is to utilise the No-SQL database because it stores the data in flexible key-value pairs in JSON documents (IIE, 2025).
- Lastly, the user experience concern can be enhanced with the above-mentioned real-time analytics.

The Types of Data to be stored in the Databases:

- The scenario mentions that the social media platform will include a variety of data types which include texts, images, videos, comments, likes, shares, meta data, cookies and more. This variety in data types warrants a big data friendly database design which harnesses the 3vs of big data.
- NoSQL is ideal for this because the traditional database struggles to handle the rapid velocity, scale, variety of data types and volume in a cost effective and efficient manner.
- Vertical scaling and hardware requirements needed to bolster traditional SQL database systems is impractical in the context of this rapidly growing big data.

Four Types of NoSQL

According to MongoDB (2023), NoSQl is an umbrella term used to classify any alternative to traditional databases. There are four types mentioned which include document databases, key-values stores, graph databases and column orientated databaseswill be discussed in this section.

Discussing the Document databases—-NoSQL type:

- An example of a document Database is MongoDB, which is used to manage noSQL data.
- Data, in document databases, is either stored in a JSON (Javascript object notation) or BSON (Binary JSON) format. BSON is used to help MongoDB or other document databases store converted JSON data into binary, which can be serialized later following retrieval.
- Document Databses are popular in application development because their flexibility allows developers to adapt their document as required to suit their applications, which is something not really practical with traditional Databases, which might require consultation with a database administrator.
- The relevance to the scenario is that the varying unstructured and semi-structured social media can be stored in this format.

Key-Value Stores:

- This is mentioned as being the simplest of NoSQL databases.
- Every element in this database is stored as a key value pair which consists of an attribute name (key) and a value, for example
- MongoDB(2023) mention that It looks like an RDBMS (Relational Database management system) with two columns.

An example of key-value stores:

Key	➡️	Value
User:45	➡️	Michaela
Post:104	➡️	“Just finished my jog”
Likes:203	➡️	78

- An example is how a dictionary is stored in C#.

Key-Value Stores in the Scenario:

- **If key values are used it would be for simple simple two row representations such as post ID and likes as the value on the social media platform.**
-

Column-Oriented Databases:

- Column-oriented databases are organised as a set of columns, and is optimised for retrieval (Fowler,n.d cited in the IIE, 2025).
- They are useful for when running analytics it can allow for the reading of columns directly without the consumption of memory and subsequent performance with unwanted data.
- Column-oriented databases differ from traditional relational databases that store entire records.
- This vertical partitioning can be traced back to the 1970s, and has been argued to optimise performance and reduce latency due to the reduced memory consumption (Abadi, Boncz and Harizopolous, 2009).

Column-Oriented Databases in the Scenario:

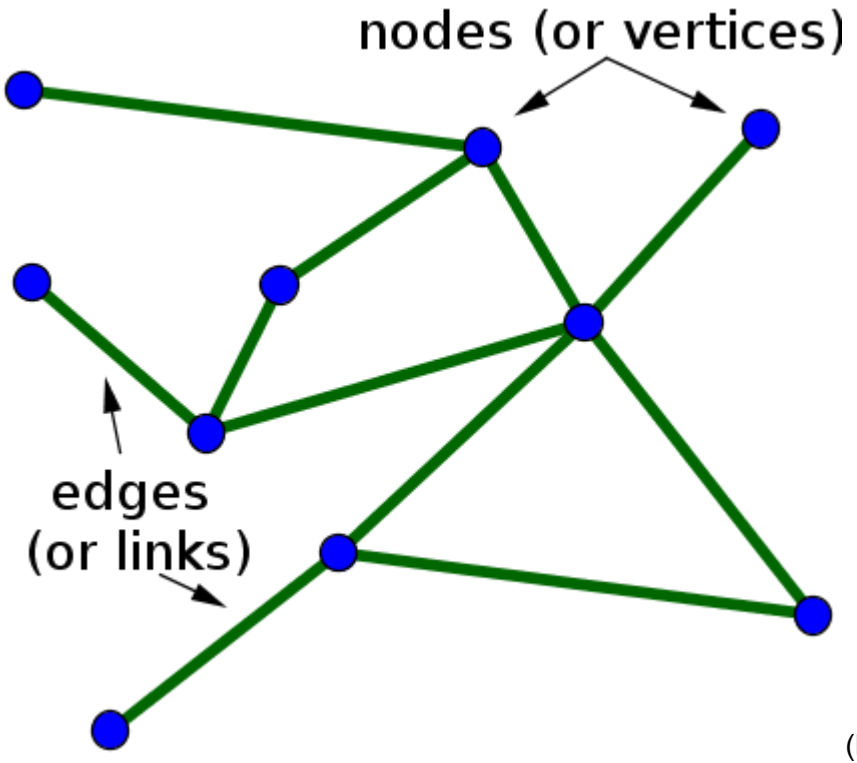
The minimal memory storage can help reduce memory strain and enhance performance and user experience on the platform.

Graph Databases:

- These are databases that focus on the relationship between data elements.
- This sees data represented as nodes, and the connections between these nodes is called links or relationships.
- Furthermore, The Graph structure is said to include nodes, edges and properties.
- These databases are schema-agnostic, which means that they allow for both unstructured and semi-structured data in a network to be stored.

Graph Databases in the Scenario:

Graph theory, through nodes can be useful for visualising large datasets on the social media, and can be useful for the analytics aspect in which node size can be used to represent interesting statistics such as demographics, gender, and other parameters of interest. This can help create narratives from the once, unstructured data-sets.



(Math Insight, 2025)

The Three Vs of Big Data
The three Vs of big data include Volume, velocity and variety. This section will be contextualising these terms with the provided scenario.

Volume: Volume refers to the large amount of data that is collected and stored.
The IIE (2025) makes reference to how 1.4 Trillion images were taken in 2020. This sheer volume and size of data is a characteristic of big data.

Volume In the scenario: there is a reference to their social media platform experiencing exponential growth in their volume of data as their users from across the globe join the platform. This necessitates strategies to make storage optimised and planned so user experience is not inhibited.

Velocity: Velocity depicts the rapid rate to which data is getting created.

Velocity in the scenario: In the contex of the scenario, the data's rapid growth in volume is intensified by the increase in users.

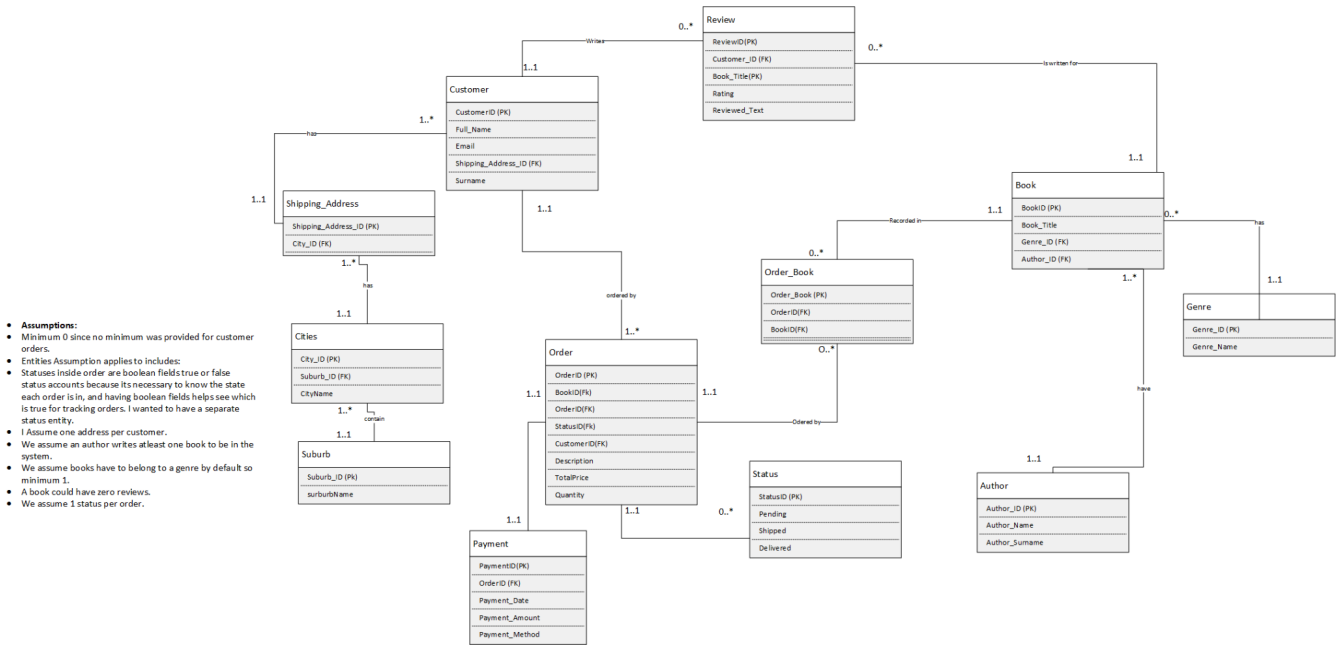
Variety: There are complex data types which necessitate flexible data structures that allow for the plethora of data types to be accommodate and stored.

Variety in the scenario:
Social Media encompasses a plethora of data types such as images, audio, video and a user can decide to post an image and others can decide to only post videos, and this variety means that the database for the social media platform must not be rigid to accommodate this variability.

Conclusion

Big data and requires unstructured and semi-structured storage to help realise its performance directives. This report argued that the social media platform's specified considerations need to harness the three v's of big data to achieve its concerns which included the complex data structures.

Question 2



Reference List:

Abadi, D.J., Boncz, P.A. and Harizopoulos, S., 2009. Column-oriented database systems. *Proceedings of the VLDB Endowment*, *2*(2), pp.1664-1665.

Exponent, 2023. SQL vs. NoSQL Explained (in 4 minutes). Accessed July 27. 2025. [Online Video]. Available: https://youtu.be/_Ss42Vb1SU4

The Independent Institute of Education (IIE), 2025. Information Stystems 1B [INSY6112 Module Outline]. The Independent Institute of Education: Unpublished.

Thomas, D. 2023. NoSQL for Real Time Analytics [online] Accessed 3 August 2025. .Available at: https://www.lonti.com/blog/nosql-for-real-time-analytics

Math Insight, 2025. Node Definition [image] Date Accessed 23 August. Accessed: https://mathinsight.org/definition/network_node.