



# INFORMATION SYSTEMS 1B

INSY6112/w

Jeyhlin Rüdwin Jaftha  
ST10472428

## Question 1

### Recommended Database Type

- When it comes to this situation, a NoSQL database is the best fit. These databases are non-relational and are built to manage all kinds of data whether it's unstructured, semi-structured, or structured. They excel at horizontal scaling and ensuring high availability. Unlike traditional relational databases, NoSQL databases don't stick to strict schemas, making them perfect for large-scale, real-time applications like social media platforms (Moniruzzaman & Hossain, 2013).

### Motivation for Using NoSQL

#### Scalability

- NoSQL databases have the ability to scale horizontally, which means you can simply add more servers to handle the growing amounts of data. This is especially important for social media platforms that have millions of users creating content every second (Sadalage & Fowler, 2013).

#### Flexibility with Complex Data Structures

- NoSQL databases can handle various formats such as JSON, BSON, and key-value pairs. This flexibility allows for the storage of text posts, photos, videos, comments, and likes without the need for a strict schema, which is a limitation of traditional relational databases (Han et al., 2011).

#### Real-Time Analytics & High Performance

- Databases like MongoDB and Cassandra are designed for low-latency queries, enabling them to update feeds and notifications almost instantly. This means users get real-time updates, which enhances engagement and overall user experience (Cattell, 2011).

### Types of Data Stored in NoSQL

- User Data → Profiles, names, bios, settings.
- Content Data → Posts, images, videos, live streams.
- Interaction Data → Likes, shares, comments, tags, mentions.
- Analytics Data → Trending hashtags, post engagement rates, location-based metrics

## Four Types of NoSQL Databases

### Document-Oriented Databases

- Store data as JSON or BSON documents. Each document can contain nested fields and varying structures, making them highly flexible.
- Example: MongoDB – widely used for handling user profiles, posts, and comments (Chodorow, 2013).

### Key-Value Stores

- Data is stored as key-value pairs. They are extremely fast and efficient for caching and session management.
- Example: Redis – ideal for storing login sessions, likes, and temporary user activity (Kleppmann, 2017).

### Column-Family Stores

- Store data in rows and dynamic columns. They are optimized for analytics and high-volume queries across distributed systems.
- Example: Apache Cassandra – perfect for storing time-series data such as likes per second or trending hashtags (Lakshman & Malik, 2010).

### Graph Databases

- Focus on relationships between entities. They are useful for modelling social connections, friendships, and recommendations.
- Example: Neo4j – used for mapping friendships, followers, and suggested friends (Robinson, Webber & Eifrem, 2015).

## **The Three Vs of Big**

### Volume

- Millions of users generate huge amounts of posts, videos, likes, and comments daily. NoSQL databases handle this scale efficiently (Chen, Mao & Liu, 2014).

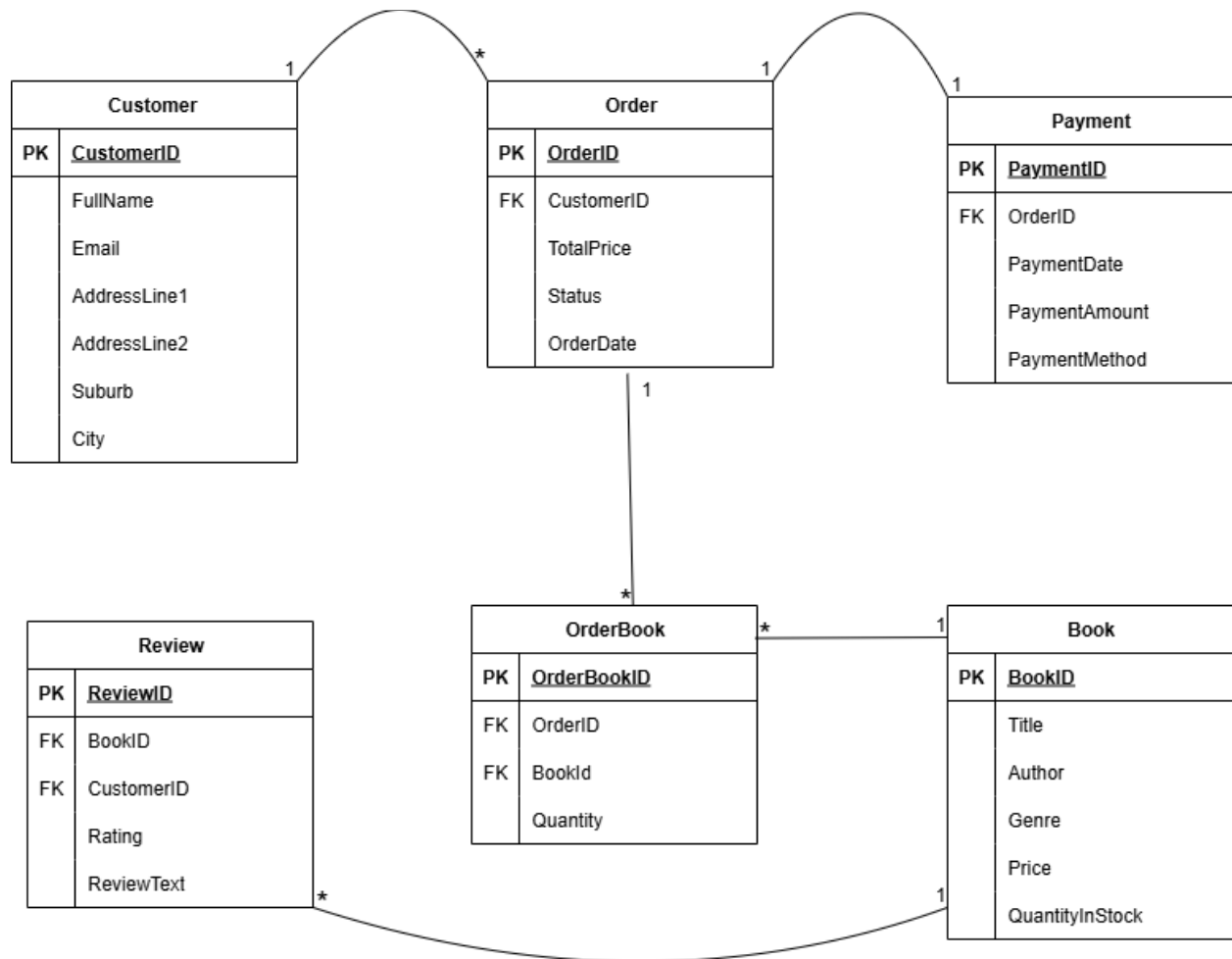
### Velocity

- Data is generated and consumed in real time (likes per second, trending hashtags). NoSQL ensures instant updates with low latency (Grolinger et al., 2013).

### Variety

- Data comes in multiple formats: text, videos, images, GIFs, and live streams. NoSQL handles unstructured and semi-structured data seamlessly (Chen, Mao & Liu, 2014).

## Question 2



## References

- Cattell, R. (2011). Scalable SQL and NoSQL data stores. *ACM SIGMOD Record*, 39(4), pp.12–27.
- Chen, M., Mao, S. & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), pp.171–209.
- Chodorow, K. (2013). *MongoDB: The Definitive Guide*. 2nd ed. O'Reilly Media.
- Grolinger, K., Higashino, W.A., Tiwari, A. & Capretz, M.A.M. (2013). Data management in cloud environments: NoSQL and NewSQL data stores. *Journal of Cloud Computing: Advances, Systems and Applications*, 2(22), pp.1–24
- Han, J., Kamber, M. & Pei, J. (2011). *Data Mining: Concepts and Techniques*. 3rd ed. Elsevier.
- Kleppmann, M. (2017). *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*. O'Reilly Media.
- Lakshman, A. & Malik, P. (2010). Cassandra: A decentralized structured storage system. *ACM SIGOPS Operating Systems Review*, 44(2), pp.35–40.
- Moniruzzaman, A.B.M. & Hossain, S.A. (2013). NoSQL database: New era of databases for big data analytics – Classification, characteristics and comparison. *International Journal of Database Theory and Application*, 6(4), pp.1–14.
- Robinson, I., Webber, J. & Eifrem, E. (2015). *Graph Databases*. 2nd ed. O'Reilly Media.
- Sadalage, P.J. & Fowler, M. (2013). *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*. Addison-Wesley.