Question 1

Summary

This report recommends the use of a NoSQL database for a large-scale social media platform. Due to the exponential data growth of the platform, the content formats available in the platform, and its requirement of real time analytics; a database system that can scale horizontally is required, that can support different data structures and can perform at high speeds. For these requirements, NoSQL databases provides the flexibility and performance, and scalability needed.

Recommendation: NoSQL

Definition:

NoSQL is a non-rational database management system that is designed to handle large sets of unstructured and semi-structured data, it supports horizontal scalability, making it perfect for real-time data processing, and supports dynamic schemas, that allows for various data formats without the need for pre-defined schemas (GeeksForGeeks, 2025).

Reasons for recommendation

Horizontal Scalability: The design of NoSQL database makes it so that it scales horizontally, this means that you add more servers to the database when needed. NoSQL databases can handle massive amounts of data by adding more servers to the database, this makes it easier to handle the increases in demand from modern day applications (GeeksForGeeks, 2023). No significant changes are needed to the underlying structure due to the nature of NoSQL databases being designed to be distributed across many servers, this makes a perfect fit for big data (Singh, 2023).

Schema Flexibility: NoSQL databases can handle unstructured and semi-structured data; this flexibility comes from the schema that is dynamic or schema-less which can accommodate dynamic changes to the data model (GeeksForGeeks, 2025), and makes it easier to store and manage a huge variety of data types like text, images and videos which is crucial for platforms where new content or content types is being generated (GeeksForGeeks, 2023).

Real-Time Performance: NoSQL databases are optimized to handle large amounts of data making them deliver a faster performance than traditional rational databases, this

is very important for applications that require real-time data access and low latency (GeeksForGeeks, 2023). NoSQL databases use various techniques to achieve faster performance like distributed architecture and in-memory storage, this ensures a rapid access of data and this could be used in applications like Uber that retrieve your location, match you with a driver and provide an estimated time of arrival all in seconds (Pingax, 2024).

Data stored in NoSQL database

User-generated content: This is data created by users themselves rather than the platform (Beverage, 2024). User-generated content includes text post that are written update, shared stories and opinions; images and videos uploaded by users which are often stored as Binary Large Objects (BLOBs) or referenced by the use od URLs; comments which is feedback and interaction with posts; and reactions that include likes, shares and other forms of engagement (Beverage, 2024).

Metadata: This provides contextual information about user interactions and content, which includes information such as users' behaviour patterns to see how users interact and navigate the platform they are on, timestamps to connection to when the content was created and last time it was interacted with, geolocation to pinpoint the location of users interacting with the platform and device information (WoopSocial. 2024). Metadata is crucial for optimizing user experience.

Social graphs: This is a representation of the network of relationships and interactions among users on a platform. These are connections such as social media's following and followers relationships that define who sees whose content and interactions through engagement with content such as liking, sharing and commenting (Webisoft, 2025). Social graphs are in use for features like friend suggestions and delivering target specific content (Webisoft, 2025).

Real-time metrics: Real-time metrics track user engagement and trends and this helps platforms adapt quickly to user preferences and optimize content strategies (Mileva, 2025). Some examples of real-time metrics include tracking engagement rates and giving a percentage of users interacting with content, track trending hashtags gaining tractions across the platform and produce live-stream viewer counts of users watching live content at any given time (Mileva, 2025).

Four types of NoSQL databases

Document Database: A document database is a type of database designed to store data in documents, usually in JASON or BSON format. Each document is a record in a

database that contains data as key-value pairs ,arrays and nested structures, this allows for complex data to be represented in a document database (Dutta, 2024). A document stores information about an object and its related data, for example a database contains three records and each give information about a user, for the first document, username is the key, and Mick is its value (Datacamp, 2022). For easy retrieval and manipulation of hierarchical data use document-oriented database like MongoDB and CouchDB (Dutta, 2024).

Key-Value Store: Key-value databases are the simplest types of NoSQL databases and store data as a collection of key-value pairs. Each unique key maps to a specific value and that can be a simple string to a complex data structure (Dutta, 2024). These databases are highly efficient for lookups and commonly used for caching, session and relational database management system and real-time data processing. Prominent examples of key-value stores are Redis and DynamoDB (Dutta, 2024).

Wide-Column Store: This is a column-orientated database which is used to store data as a collection of columns, where each column is treated differently (Datacamp, 2022). This type of database is designed to read and retrieve data more efficiently and with greater speed (GeeksForGeeks *b*, 2025). Columns can be added on the fly for an evolving schema and columns can be grouped into column families, which are groups of related columns. Wide- column databases are good for storing massive amounts of data, its easily scalable across many machines and is very good for handling many write operations. Wide-column database examples are Apache Casandra or Azure's CosmosDB (Simonyan, 2024).

Graph Databases: A graph database uses graph structures to represent and store data (Dutta, 2024). This type of database is used to store, map and search relationships between nodes through edges, nodes represent data elements also called an entity (Datacamp, 2022). Managing and querying complex data relationships is where graph databases excel in and that makes them perfect for applications like social networks and fraud detection (Dutta, 2024). Amazon and Meta use graph databases to recommend products, and represent and analyse connections between social media accounts. Some examples of graph databases are Neo4j, Amazon Neptune and Dgraph (Simonyan, 2024).

Three V's of big data

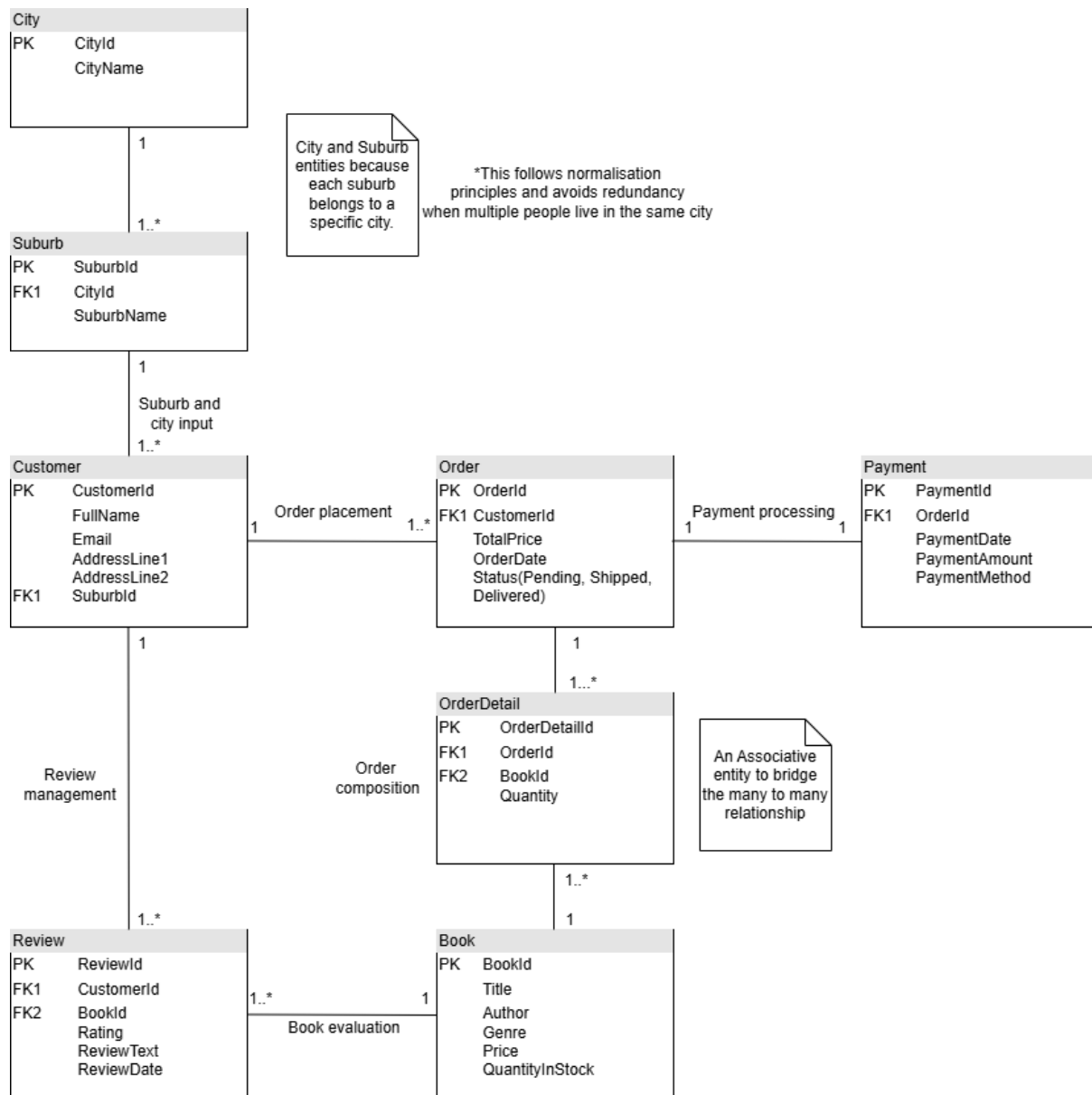Volume: This refers to huge amounts of any kind of data from many different sources (Romero, Ventura, and Herrera, 2016), the social media platform connects millions of users worldwide and generates an astronomical amount of data. Since there is constant interactions with the platform, new users join and new features get introduced to the platform the volume of data generated from post, comments and interactions

grows rapidly. The amount of data generated poses a major challenge in accommodating the growth but benefits of gathering, processing and analysing the large amounts of data generated is obtaining valuable knowledge (Romero, et al., 2016).

Variety: This refers to gathering different types of data from different mediums like smartphones or social networks (Romero, et al., 2016), this platform is the perfect example as it gathers all types of data from users who interact with it such as photos, videos, text posts and much more. The data gathered can be in a structured or unstructured format and since the platform uses NoSQL it puts data in an unstructured format making it perfect for the platform's varied data types (Romero, et al., 2016).

Velocity: This refers to the speed of data transfers, data's contents are constantly changing in many ways through the absorption of complementary data, legacy data and all sorts of streamed data from multiple sources (Romero, et al., 2016). The speed of data transfers can make or break user experience, users expect quick and instants updates on theirs feed when interacting with a platform to avoid poor management caused by delays due to the speed of data transfer new methods and algorithms are need to better and adequately process and analyse data (Romero, et al., 2016).

Question 2

City

PK	CityId
	CityName

City and Suburb entities because each suburb belongs to a specific city.

*This follows normalisation principles and avoids redundancy when multiple people live in the same city

1

1..*

Suburb

PK	SuburbId
FK1	CityId
	SuburbName

1

Suburb and city input

1..*

Customer

PK	CustomerId
	FullName
	Email
	AddressLine1
	AddressLine2
FK1	SuburbId

1 — Order placement — 1..*

Order

PK	OrderId
FK1	CustomerId
	TotalPrice
	OrderDate
	Status(Pending, Shipped, Delivered)

1 — Payment processing — 1

Payment

PK	PaymentId
FK1	OrderId
	PaymentDate
	PaymentAmount
	PaymentMethod

1

1...*

OrderDetail

PK	OrderDetailId
FK1	OrderId
FK2	BookId
	Quantity

An Associative entity to bridge the many to many relationship

Order composition

Review management

1

1..*

1..*

1

Review

PK	ReviewId
FK1	CustomerId
FK2	BookId
	Rating
	ReviewText
	ReviewDate

1..* — Book evaluation — 1

Book

PK	BookId
	Title
	Author
	Genre
	Price
	QuantityInStock

References

GeeksForGeeks. 2025. Introduction to NoSQL, [Online] 12 March. Available at: <Introduction to NoSQL - GeeksforGeeks> [Accessed 24 March 2025].

GeeksForGeeks. 2023. Top 5 reasons to choose NoSQL, [Online] 20 February. Available at: <Top 5 Reasons to Choose NoSQL - GeeksforGeeks> [Accessed 24 March 2025].

Singh, A. 2023. Exploring NoSQL databases: When and Why to use them, Medium, [Online] 2 June. Available at: <Exploring NoSQL Databases: When and Why to Use Them | by Ankush Singh | Medium> [Accessed 24 March 2025].

Dutta, S. 2024. What is a NoSQL Database: Understanding the Evolution of Data Management, Sprinkle, [Online] 4 June. Available at: <Understanding NoSQL Databases: Types, Advantages, and Use Cases | Comprehensive Guide> [Accessed 24 March 2025].

Datacamp. 2022. NoSQL Databases: What Every Data Scientist Needs to Know, [Online] 24 June. Available at: <NoSQL Databases - Types of NoSQL Databases and How to Use Them | DataCamp> [Accessed 24 March 2025].

GeeksForGeeks *b*. 2025. Types of NoSQL databases, [Online] 25 February. Available at: < Types of NoSQL Databases - GeeksforGeeks> [Accessed 24 March 2025].

Pingax. 2024. Managing NoSQL Databases for Real-Time Applications, [Online]. Available at: <Managing NoSQL Databases for Real-Time Applications - Pingax> [Accessed 24 March 2025].

Beverage, C. 2024. User-Generated Content (UGC): What it is + why it matters, Hootsuite, [Blog] 26 March. Available at: < What is User-Generated Content? And Why is it Important?> [Accessed 25 March].

WoopSocial. 2024. Metadata, [Blog]. Available at: < What is Metadata? Definition & Meaning - WoopSocial Social Media Glossary> [Accessed 25 March 2025].

Webisoft. 2025. The Social Graph: Understanding the Web of Connections, [Blog] 8 February. Available at: < The Social Graph: Understanding the Web of Connections - Webisoft Blog> [Accessed 25 March].

Mileva, G. 2025. Real-Time Social Media Analytics Tools for Your Business Intelligence Arsenal, [Online]. Available at: < Real-Time Social Media Analytics Tools for Your Business Intelligence Arsenal> [Accessed 25 March].

Simonyan, H. 2024. NoSQL Databases Explained: All Types & When to Use Them, Substack, [Blog] 30 April. Available at: < NoSQL Databases Explained: All Types & When to Use Them> [Accessed 25 March].

Romero, V., Ventura, S. and Herrera, F. 2016. An Overview on Social Big Data: Definition, Data Sources, and Challenges, International Journal of Interactive Multimedia and Artificial Intelligence,[e-journal] 3(4), p. 57. <Social big data: Recent achievements and new challenges - PMC>[Accessed 25 March].