Information System 1B Databases

Assignment 1

ST10439133 Camryn Naidoo

**QUESTION 1:**

Report on Database Recommendation for Social Media Platform

NoSQL databases are usually the recommended remedy to handle huge data sets that may not be completely accommodated within traditional relational databases. They are specifically designed for the handling of unstructured, semi-structured, or dynamic forms of data and ensure scalability, flexibility, and optimization. With NoSQL databases, horizontal scaling can be achieved with the possibility of adding extra servers to handle growing data. These databases are used widely in applications such as social networking sites, e-commerce websites, and real-time analytics with the requirement of fast operations (DataStax, 2024).

Reasons for Recommending NoSQL Database:

* Scalability

One of the greatest advantages of these databases is their horizontal scalability. This is particularly crucial for social media websites when more people become members, the data increases very quickly. Relational databases do not perform well in such settings since they are vertically scalable and hence more powerful servers are needed for optimal performance. As the platform grows, NoSQL databases offer seamless performance by distributing data across multiple machines (DataStax, 2024).

* Flexible Data Structures

Social media sites create a big variety of content (e.g., text posts, images, video, likes, shares, comments), and such a dynamic content storage capability is required. NoSQL databases either don't have a plan or feature as flexible representation so that they can store various types of content. Relational databases, on the other hand, require a fixed schema so they are not as suitable for storing variable, dynamic types of content structure. NoSQL databases like MongoDB store data as documents (JSON-like structures), allowing easy accommodation of different content types without restructuring (GeeksforGeeks, 2024).

- Real-Time Analytics

Social media platforms rely on real-time analytics to know trending subjects, most engaged posts, and the level of engagement. NoSQL databases like Cassandra or Elasticsearch are optimized for low potential data processing and are capable of handling high-speed reads and writes efficiently. Such databases work optimally in applications where data keeps changing, for example, user interaction with a post. Relational databases would struggle to achieve the real-time responsiveness needed in such applications (Curate Partners, 2023).

Data Types Stored in NoSQL Databases:

- Text Posts: Plain text content like messages, reviews, or status updates that are posted by users.

- Multimedia Content: Photos and videos that users post. Because the data types vary in size and structure, they require flexible storage facilities.

- User Interactions: Information regarding how users interact with posts, e.g., comments, likes and shares. The information must be updated constantly because it is usually highly active.

-Statistics that are used to track user engagement, e.g., the number of views a post gets or the number of likes and comments per minute, are referred to as real-time metrics. This data needs fast and high-scale processing.

Types of NoSQL Databases:

* Document-based databases (such as MongoDB) Data is stored in JSON-like documents, which may have varying structures, in a document-based NoSQL database. Applications with dynamically fluctuating data structures, like user posts, comments, and multimedia content on a social network, are greatly served by its adaptability. Because of its documented capability to insert and retrieve documents quickly, MongoDB is well suited for high data velocity environments. Furthermore, it supports complex queries without compromising performance because of its ability to index nested structures (GeeksforGeeks, 2024).
* High-value stores (e.g., Redis) Data is kept in key-value pairs within key-value stores, which are among the simplest types of NoSQL databases. An example is Redis, which is well known for its high read and write speeds. In order to make real-time updates, like notifications, likes, and comments, available in an instant, this kind of database is especially useful in caching interactions from users or session data on social networking websites. Key-value stores are suited best for frequently accessed data with little overhead as they are speedy and convenient (PhoenixNAP, 2024).
* Graph databases (like Neo4j) Graph databases employ edges to reflect entities and their interactions. Neo4j is a fine example of this category. Graph databases are specially helpful for analyzing and managing intricate user relationships on social media like friendships, followers, and activities like likes, comments, and shares. They are particularly apt for functionality like buddy recommendations or content suggestion algorithms since they allow for rapid exploration of these relationships (GeeksforGeeks, 2024).

The Three Vs of Big Data:

* Volume

The amount of data generated by a social media platform is huge and growing at an explosive rate. Each user posts, comments, likes, shares, and other interactions, leading to a large dataset. Databases like Cassandra, the NoSQL database, are designed particularly to handle this huge amount by duplicating data on several servers, such that scalability is not disadvantaged and performance is not impacted. Horizontally scaling is crucial in handling voluminous data (DataStax, 2024).

* Velocity

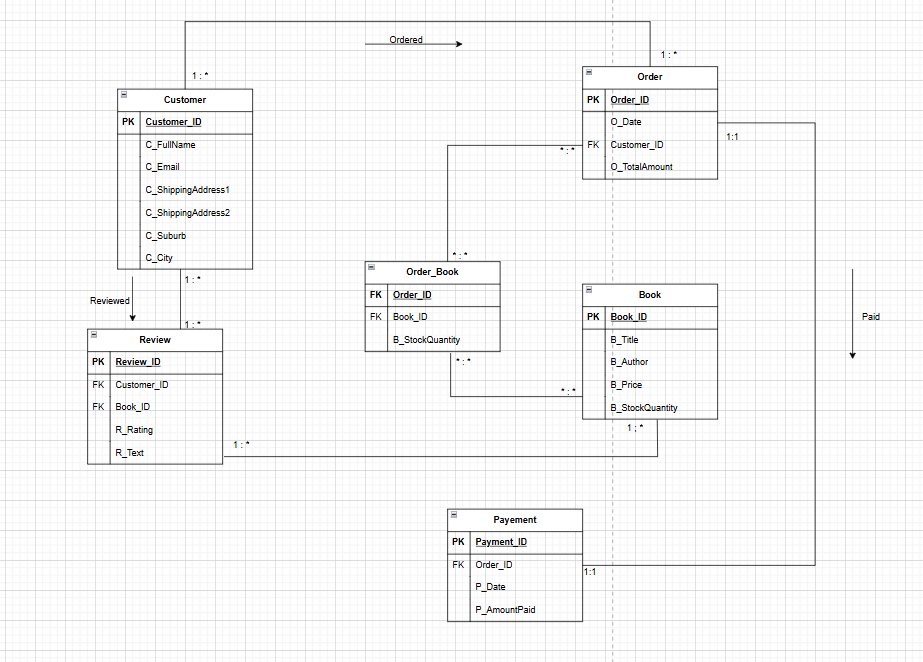
Velocity is the speed at which data is being generated, processed, and examined. On a social media platform, new content is uploaded and interacted with in real-time, hence the need for real-time updates on user feeds and analytics. NoSQL databases are suited for high-velocity data as they are capable of inserting, retrieving, and processing data quickly in real-time. For example, Redis has rapid access to up-to-date highly updated data, sending timely notifications and feed updates (LinkedIn, 2023).

* Variety

Variety of data on a social media platform includes text posts, images, videos, comments and likes, each with its own format and structure. NoSQL databases, especially MongoDB, are well-suited to handle such diverse forms of data as they do not have a rigid representation. This allows it to easily incorporate new forms of content and changing user behaviours without having to redesign the database schema (DataStax, 2024).

**QUESTION 2:**

ONLINE BOOK STORE SYSTEM ENTITY RELATION DIAGRAM

****

**BIBLIOGRAPHY:**

1. Curate Partners (2023) Unlocking the Power of HBase: Distributed NoSQL for Scalable Real-Time Data Solutions. Available at: <https://curatepartners.com/blogs/skills-tools-platforms/unlocking-the-power-of-hbase-distributed-nosql-for-scalable-real-time-data-solutions> (Accessed: 25 March 2025).
2. DataStax (2024) NoSQL Use Cases: Social Media and Beyond. Available at: <https://www.datastax.com/guides/nosql-use-cases> (Accessed: 25 March 2025).
3. GeeksforGeeks (2024) Types of NoSQL Databases. Available at: <https://www.geeksforgeeks.org/types-of-nosql-databases> (Accessed: 25 March 2025).
4. LinkedIn (2023) NoSQL Databases: The Future of Scalable and Flexible Data. Available at: <https://www.linkedin.com/pulse/nosql-databases-future-scalable-flexible-data-nithya-joseph-hdd7c> (Accessed: 25 March 2025).
5. PhoenixNAP (2024) NoSQL Database Types and Their Use Cases. Available at: <https://phoenixnap.com/kb/nosql-database-types> (Accessed: 25 March 2025).