INSY16112
Assignment 1

Database Recommendation Report for Social Media Platform

ETHAN GUY ALGEO – ST10444972

Question 1)

Recommended Database Type: NoSQL Database

Definition

A NoSQL (Not Only SQL) database is a non-relational database design that provides a mechanism for storage and retrieval of data that is modelled in means other than the tabular relations used in relational databases [1]. It offers flexible schemas, horizontal scalability, and high performance for large-scale distributed data systems.

Motivation for NoSQL Database

1. **Horizontal Scalability**

   NoSQL databases are designed to scale out horizontally across multiple servers, which is crucial for a social media platform experiencing exponential data growth. Unlike traditional relational databases that scale vertically, NoSQL databases can easily add more servers to handle increased data volume and user interactions without significant performance degradation.

2. **Flexible Data Modelling**

   The complex and varied data structures in a social media platform (text posts, images, videos, comments) require a database that can handle diverse and evolving data types. NoSQL databases support schemeless designs, allowing different types of content to be stored without rigid schema constraints. This flexibility enables quick adaptation to new content types like live streaming or virtual reality posts.

3. **High-Performance Real-Time Processing**

   NoSQL databases excel at real-time data processing and analytics. They can handle massive write and read operations simultaneously, which is essential for updating trending topics, engagement metrics, and user feeds almost instantly. The distributed architecture allows for faster data retrieval and processing compared to traditional relational databases.

Types of Data to be Stored

1. User profile information

2. Content metadata (posts, comments, likes)

3. Media files (images, videos)

4. User interaction logs

5. Real-time analytics and engagement metrics

6. Recommendation and personalization data

7. Notification and activity streams

Four Types of NoSQL Databases

1. **Document Store (e.g., MongoDB)** A document store database stores data in flexible, JSON-like documents called BSON (Binary JSON). Each document can have a different structure, making it ideal for storing varied content types like user posts, comments, and profiles. Documents are organized into collections, which are analogous to tables in relational databases but offer much more flexibility in data representation.

2. **Key-Value Store (e.g., Redis)** Key-value stores are the simplest NoSQL database type, where each item is stored as an attribute name (key) together with its value. They are extremely fast and efficient for caching, storing session information, and managing real-time data like user notifications, online status, and temporary computational results. Redis, for example, can handle millions of read/write operations per second.

3. **Column-Family Store (e.g., Cassandra)** Column-family stores organize data into column families, which are containers for rows [2] They are excellent for time-series data and can handle massive write loads with high availability. In a social media context, they can efficiently store and retrieve large volumes of user interaction data, such as likes, shares, and engagement metrics across different time periods.

4. **Graph Database (e.g., Neo4j)** Graph databases store data in nodes and edges, making them perfect for representing complex relationships between users, content, and interactions. They excel at social network analysis, recommendation engines, and tracking user connections. Graph databases can quickly traverse relationships to suggest friends, recommend content, or analyze network structures.

The Three Vs of Big Data in Social Media Context

1. **Volume**

The massive scale of data generated by millions of users creates an enormous volume challenge. Every post, comment, like, and interaction generates data, leading to exponential growth that requires a database capable of handling petabytes of information efficiently.

2. **Velocity**

The speed at which data is generated and processed is critical. User interactions happen in real-time, requiring instant updates to feeds, notifications, and analytics. The database must support high-speed data ingestion and processing to maintain a responsive user experience.

3. **Variety**

Social media platforms deal with diverse data types and structures. From text posts and images to videos and live streams, the data variety is extensive. The database must accommodate these different formats without compromising performance or requiring complex preprocessing.
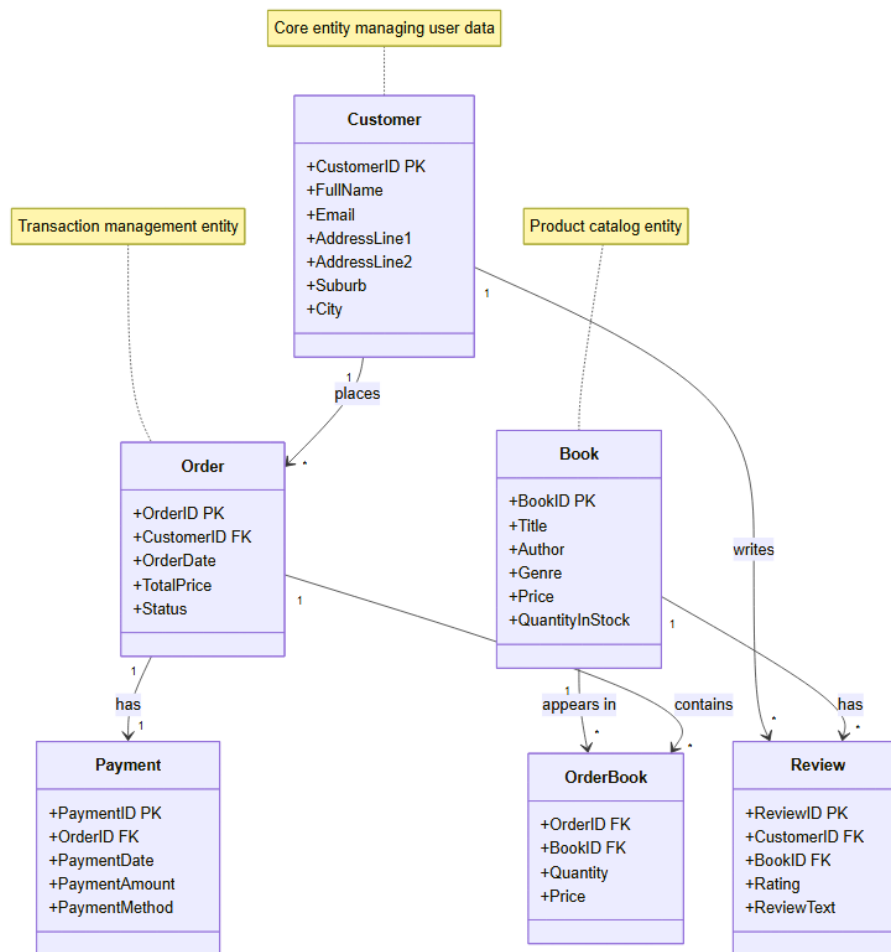
[3]

Conclusion

A NoSQL database provides the most suitable solution for this large-scale social media platform, offering the flexibility, scalability, and performance required to handle complex, high-volume, and rapidly changing data environments.

Question 2)

Online Bookstore ERD

Core entity managing user data

Customer

+CustomerID PK
+FullName
+Email
+AddressLine1
+AddressLine2
+Suburb
+City

Transaction management entity

Product catalog entity

1

1   places

Order

+OrderID PK
+CustomerID FK
+OrderDate
+TotalPrice
+Status

Book

+BookID PK
+Title
+Author
+Genre
+Price
+QuantityInStock

writes

1

1

1   has

1

appears in   contains   has

1

Payment

+PaymentID PK
+OrderID FK
+PaymentDate
+PaymentAmount
+PaymentMethod

OrderBook

+OrderID FK
+BookID FK
+Quantity
+Price

Review

+ReviewID PK
+CustomerID FK
+BookID FK
+Rating
+ReviewText

References

[1] R. Cattel, "Scalable SQL and NoSQL data stores," Special Interest Group on Management of Data, 2011.

[2] A. Lakshman and P. Malik, "Cassandra: a decentralized structured storage system," AGM SIGOPS, 2010.

[3] A. B. Moko and P. Asagba, "Big Data and NoSQL Databases Architecture: A Review," 2021.