INSY6112 1B

Assignment 1

Kaedon Naidoo ST10293141

Question 1:

Recommendation Report: Database Selection for a Large Social Media Platform

1. Suggested Database Type:

The suggested type of database for the social media platform is a NoSQL database. NoSQL databases are built to efficiently manage large amounts of unstructured and semi-structured data. They offer flexible schema designs, support for horizontal scaling, and high-performance real-time data processing, making them well-suited for dynamic and data-heavy applications like social media platforms. (Nguyen, 2024)

2. Justification for Choosing NoSQL:

2.1 Scalability:

NoSQL databases exhibit excellent scalability and facilitate horizontal scaling, meaning that extra servers can be added to distribute the workload. This capability is essential for a social media platform, where the amount of data can grow exponentially as the user base expands. (Nguyen, 2024)

2.2 Adaptable Data Structures:

In contrast to traditional relational databases, NoSQL databases do not impose a strict schema, which allows for the accommodation of diverse content types, including text, images, videos, and user interactions. This adaptability is crucial for platforms that regularly add new features and data formats. (Nguyen, 2024)

2.3 Immediate Data Processing:

NoSQL databases facilitate real-time analytics by allowing rapid read/write operations. Social media platforms need real-time updates for feeds, notifications, and trending topics, which can be effectively managed through NoSQL's distributed architecture. (Nguyen, 2024)

3. Categories of Data Maintained in the Database:

User Accounts: Details such as usernames, email addresses, profile images, and user settings. (Nguyen, 2024)

Content and Posts: Written posts, photographs, videos, live broadcasts, and virtual reality content shared by users. (Nguyen, 2024)

User Engagement: Likes, comments, shares, reactions, and metrics related to user interaction. (Nguyen, 2024)

Alerts: Immediate notifications regarding mentions, likes, new followers, and other updates. (Nguyen, 2024)

Analytical Information: Popular topics, engagement statistics, and various performance metrics gathered for insights and recommendations. (Nguyen, 2024)

4. Types of NoSQL Databases:

4.1 Document-Oriented Database:

A document database organizes data in documents similar to JSON, which is advantageous for managing complex and hierarchical data formats. It permits a flexible schema design, allowing different types of content like text, images, and metadata to be stored within a single document. Notable examples are MongoDB and CouchDB. (Jain, 2025)

4.2 Key-Value Store:

A key-value store is a straightforward and efficient type of database that retains data as pairs of keys and values. It is highly optimized for rapid lookups and caching, making it well-suited for storing user sessions, preferences, and commonly accessed information. Examples include Redis and Amazon DynamoDB. (Jain, 2025)

4.3 Column-Family Store:

Column-family databases organize data in columns instead of rows, which facilitates the efficient retrieval of extensive structured data. They are especially beneficial for storing user analytics, engagement metrics, and timeline feeds. Notable examples include Apache Cassandra and HBase. (Jain, 2025)

4.4 Graph Database:

Graph databases are specifically built to efficiently store and query the relationships between entities. They excel in modelling social connections, friend networks, and recommendation systems based on user interactions. Examples of such databases include Neo4j and Amazon Neptune. (Jain, 2025)

5. Big Data's Three Vs in the Given Scenario:

5.1 Volume:

Every day, the social networking site produces a vast amount of data, including uploads of videos, likes, comments, and posts. This enormous volume of data can be handled by NoSQL databases by spreading it across several servers. (Jain, 2025)

5.2 Velocity:

New data is being created and processed at a very rapid rate. NoSQL is the best option since real-time analytics, alerts, and fast feed changes demand a database that can process data quickly. (Jain, 2025)

5.3 Variety:

The platform works with a variety of data formats, such as semi-structured (JSON documents), unstructured (videos and photos), and structured (user profiles). Because NoSQL databases allow a variety of data models and formats, they effectively manage this diversity. (Jain, 2025)

In Conclusion:

A NoSQL database is the ideal option due to its high scalability, flexible schema, and real-time processing capabilities, which are all necessary for the social networking platform. Document-oriented, key-value, column-family, and graph databases are among the NoSQL varieties that provide specific solutions for certain facets of the platform's data management requirements.

Question 2:

The following is a text presentation of the Entity Relationship Diagram (ERD) in UML notation:

Entities with Attributes:

Customer

CustomerID (PK)

FullName

Email

AddressLine1

AddressLine2

Suburb

City

Book

BookID (PK)

Title

Author

Genre

Price

QuantityInStock

Order

OrderID (PK)

TotalPrice

Status (Pending/Shipped/Delivered)

OrderDate

Payment

PaymentID (PK)

PaymentDate

Amount

PaymentMethod

Review

ReviewID (PK)

Rating (1-5)

ReviewText

ReviewDate

OrderBook (Junction table for Order-Book many-to-many)

Quantity

Relationships with Multiplicities:

Customer to Order:

One-to-Many (1..*)

A customer can place many orders

Each order belongs to exactly one customer

Customer to Review:

One-to-Many (1..*)

A customer can write many reviews

Each review is written by exactly one customer

Book to Review:

One-to-Many (1..*)

A book can have many reviews

Each review is for exactly one book

Order to Payment:

One-to-One (1-1)

Each order has exactly one payment

Each payment belongs to exactly one order

Order to Book (through OrderBook):

Many-to-Many (resolved via junction table)

An order can contain many books (with quantities)

A book can appear in many orders

Business Rules Enforced:

Customer Constraints:

Unique CustomerID

Complete address information

Can have multiple orders

Book Constraints:

Unique BookID

Single author per book

Multiple books per genre

Order Constraints:

Must contain at least one book

Calculated total price

Specific status values

Payment Constraints:

Must be associated with an order

Multiple payments can use same method

Review Constraints:

One review per book per customer

Rating between 1-5 stars

Primary Keys:

Customer: CustomerID

Book: BookID

Order: OrderID

Payment: PaymentID

Review: ReviewID
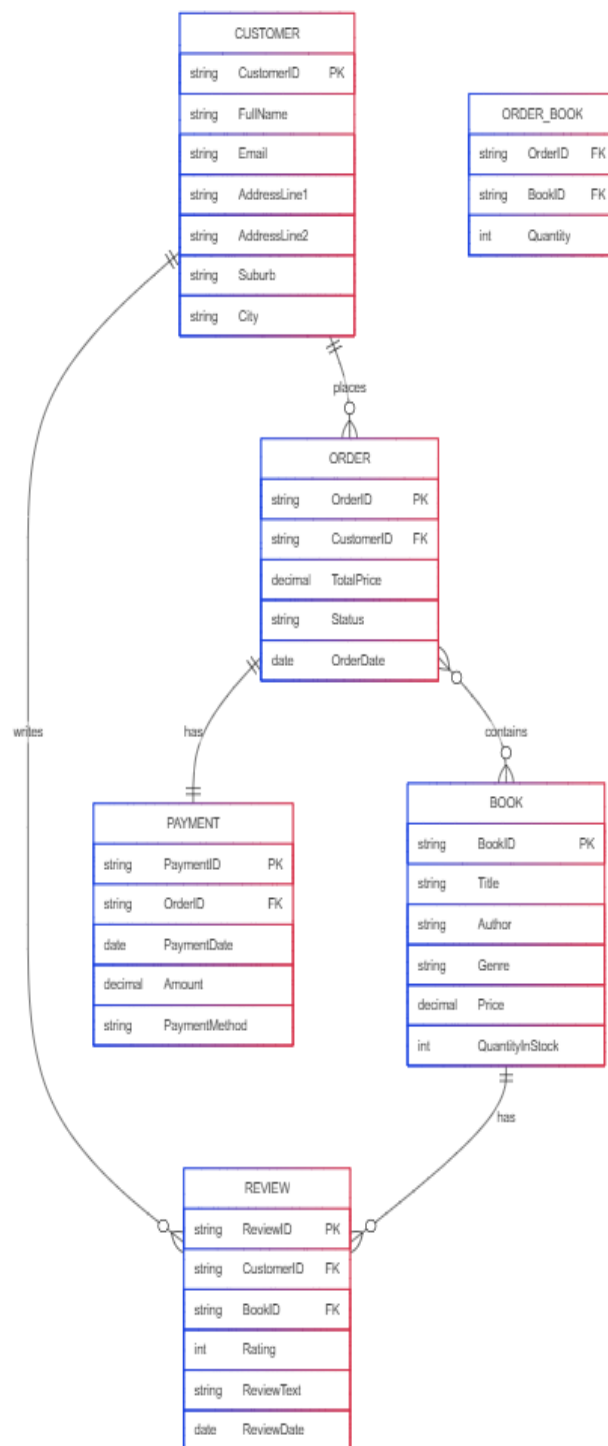
Foreign Keys:

Order: CustomerID references Customer

Payment: OrderID references Order

Review: CustomerID references Customer, BookID references Book

OrderBook: OrderID references Order, BookID references Book

Entities, attributes, relationships, multiplicities, and business rules in standard UML notation format are all included in this textual representation, just like in a graphical ERD.

The following is the diagram:

CUSTOMER

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

ORDER_BOOK

string	OrderID	FK
string	BookID	FK
int	Quantity	

places

ORDER

string	OrderID	PK
string	CustomerID	FK
decimal	TotalPrice	
string	Status	
date	OrderDate	

writes
has
contains

PAYMENT

string	PaymentID	PK
string	OrderID	FK
date	PaymentDate	
decimal	Amount	
string	PaymentMethod	

BOOK

string	BookID	PK
string	Title	
string	Author	
string	Genre	
decimal	Price	
int	QuantityInStock	

has

REVIEW

string	ReviewID	PK
string	CustomerID	FK
string	BookID	FK
int	Rating	
string	ReviewText	
date	ReviewDate	

Draw.io/ (Accessed: 24 March 2025).

Reference List:

Diagram:

Free Flowchart Maker and diagrams online (2025) Flowchart Maker & Online Diagram

Software. Available at: http://Draw.io/ (Accessed: 24 March 2025).

altexsoft:

altexsoft, software r&d engineering. (n.d.). NoSQL Databases Visually Explained with Examples. altexsoft.

https://www.altexsoft.com/blog/nosql-databases/

Nguyen, L. (2024). altexsoft: NoSQL Databases Visually Explained with Examples.

geeksforgeeks:

geeksforgeeks. (n.d.). geeksforgeeks documentation. geeksforgeeks.

https://www.geeksforgeeks.org/types-of-nosql-databases/

Jain, S. (2025). geeksforgeeks: Types of NoSQL Databases.