INFORMATION SYSTEMS 1B

chrisma dibwe

ST10443058   VARSITY COLLEGE

QUESTION 1

Database Recommendation Report

I believe a large scale social media platform as described in the text would require a high-performance database that is capable of handling massive data volumes , real time updates and complex data structures. In my report i will evaluate the feasibility of RELATIONAL (SQL) vs. NoSQL databases and recommends the optimal solution based on scalability , flexibility , and real-time performance.

Introduction

Modern social media platforms like for instance , instagram , tiktok etc they process billions of interaction per second, this includes texts , posts , videos , image , likes , shares , comments , even reposts. The problem with traditional relational databases (RDBMS) is that it struggles with the high data variety and velocity required in real-time applications.

Ill evaluate the best database model for this scenario by considering
- The nature of the data (structured vs. unstructured)
- The scalability needs of a rapidly growing platform
- The real-time analytics capabilities required for user engagement

Recommended Database Type

3.1 I recommend the database : NoSQL

I recommend this due to its ability to:
- Handle massive volumes of unstructured data from user interactions
- Scale horizontally (adding more servers instead of increasing observer capacity)
- Provide real time data processing , ensuring instant updates for likes , comments and shares

So unlike SQL databases , which solely rely on fixed schemas , NoSQL supports document-based , key-value , column- family and graph models , therefor offering greater flexibility for evolving media features (Stonebraker, 2005)

Justification for NoSQL

Scalability and High Data Volume

The issue , now the traditional RDBMS scales vertically (by upgrading hardware) , which can be quite costly and Limited

The solution to this is NoSQL databases use horizontal scaling (adding servers dynamically) to be able to manage millions of concurrent users

For instance Facebooks database stores over 300 petabytes of user-generated content , handled by Apache Cassandra (NoSQL)

(Lakshman, 2010)

Flexible Data Handling

The issue is that social media content is extremely diverse , theres texts , images , videos , VR reactions.
The solution to this NoSQL supports schema-less structures , which means that we can introduce new data types with in any way redesigning the database , a great example of this is how instagram switched from PostgreSQL to Cassandra(NoSQL) data to be able to better support their photo sharing scale (Lakshman, 2010) (Grolinger, 2014)

Real-Time Analytics and Performance

New the common issue is that SQL databases rely solely on complex joins that slow performance , the solution is NoSQL's distributed architecture allows for real time updates for notifications , trending posts and engagement metrics. A great example of this is twitter , twitter's real-time analytics system relies on Apache storm (NoSQL-based) to allow them to process hundreds of thousands of tweets per second (Lakshman, 2010)

The types of data stored

- USER PROFILES

This includes : Names of people , email addresses , Avatars , their follower lists and lastly their following lists
- POSTS AND MEDIA

This includes: Text , Images , VIdeos , live streams , and even GIFs (graphics interchange format)
- USER INTERACTIONS

This includes : likes , shares , Comments , story views
- NOTIFICATIONS

This includes : Tags , Mentions , in terms of facebook friend requests , following requests , message requests
- ENGAGEMENT

This includes : Trending , Hashtags , impressions
- DATA

Clicks (Grolinger, 2014)

Types of NoSQL Databases

- Document-based

Example : MongoDB
This is used for storing posts , comments and profiles . this database stores data in flexible , JSON-Like documents instead of traditional tables. Its perfect for handling social media posts , comments and user profiles since ech document can store different fields without needing a structure. Platforms make use of document based databases to manage user-generated content effienctly
- Key-Value

Example: Redis
This is primarily used for fast lookups for notification and session data . they are mainly used for handling real-time notifications , session management and caching data . when you get instant notications , a key-value database like redis is likely behind it

- Column- family

Example: Trending topics , Analytics and recommendations . These databases store data in columns rather than rows, making them great for handling large-scale analytics. They're often   used for tracking trending topics, engagement insights, and recommendation systems. When you see a "Trending" section or get personalized content suggestions, a column-family database is working behind the scenes.

- Graph database

Example : Neo4j
This is commonly used for storing relationships like followers , likes and shares . they haelp platforms like linkedin or facebook track connections between users , suggest new friends and analyze interactions. (Cattell, 2011)

The 3 Vs of Big Data

- VOLUME

This refers to large quantities of data.
Application to social media : Billions of likes , posts , shares daily . platforms like instagram , tiktok etc handle an extreme amount of data which means that they need powerful systems to store and process everything without crashing.

- VELOCITY

This refers to the speed of processing , how fast data is created and processed.
Application to Social media: instant updates and real time management tracking

- VARIETY

This refers to different types of data.
Application to social media : this would refer to text , images , videos , VR Content. Since people create content in so many formats , platforms have to be able to handle everything from a simple tweet to a high-quality 4k video without slowing down. (Laney, 2001)

Conclusion

I believe a NoSQL database is the ideal solution for a large social media platform due to its High scalability meaning that it supports billions of interactions , secondly i recommended it because its Schema flexibility which means that it can handle unstructured content and lastly Real time performance meaning it delivers instant updates . Global platforms like facebook , twitter and instagram , NoSQL databases(MongoDB , Cassandra , Redis, Neo4j) have been shown to be essential in handling big data demands
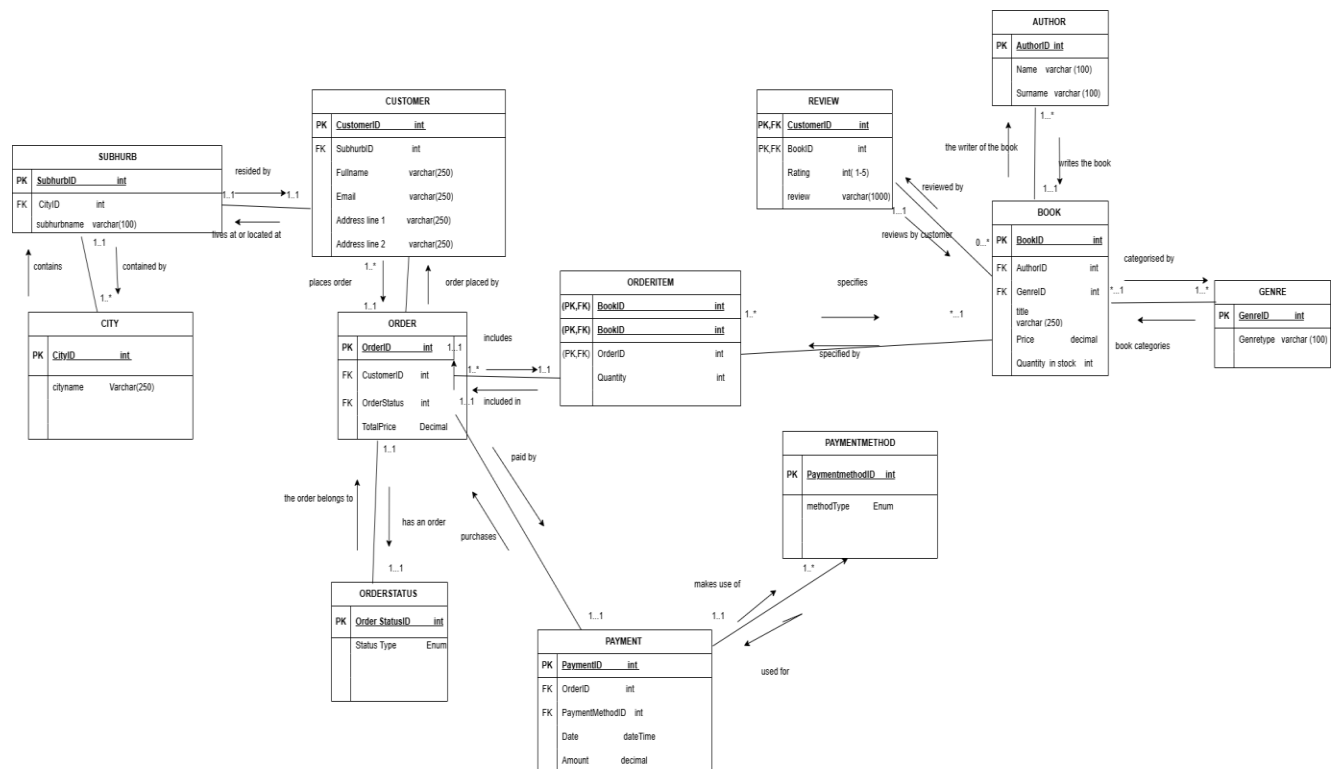
QUESTION 2

SUBHURB

PK	SubhurbID	int
FK	CityID	int
	subhurbname	varchar(100)

CUSTOMER

PK	CustomerID	int
FK	SubhurbID	int
	Fullname	varchar(250)
	Email	varchar(250)
	Address line 1	varchar(250)
	Address line 2	varchar(250)

CITY

PK	CityID	int
	cityname	Varchar(250)

ORDER

PK	OrderID	int
FK	CustomerID	int
FK	OrderStatus	int
	TotalPrice	Decimal

ORDERITEM

(PK,FK)	BookID	int
(PK,FK)	BookID	int
(PK,FK)	OrderID	int
	Quantity	int

REVIEW

PK,FK	CustomerID	int
PK,FK	BookID	int
	Rating	int(1-5)
	review	varchar(1000)

AUTHOR

PK	AuthorID int
	Name varchar (100)
	Surname varchar (100)

BOOK

PK	BookID	int
FK	AuthorID	int
FK	GenreID	int
	title varchar (250)	
	Price	decimal
	Quantity in stock	int

GENRE

PK	GenreID	int
	Genretype	varchar (100)

PAYMENTMETHOD

PK	PaymentmethodID	int
	methodType	Enum

ORDERSTATUS

PK	Order StatusID	int
	Status Type	Enum

PAYMENT

PK	PaymentID	int
FK	OrderID	int
FK	PaymentMethodID	int
	Date	dateTime
	Amount	decimal

Figure 1 UML

References

Cattell, R., 2011. *'Scalable SQL and NoSQL Data Stores', ACM Computing Surveys.* s.l.:s.n.

Grolinger, K. e. a., 2014. *Challenges for MapReduce in Big Data Processing. IEEE Cloud Computing..* s.l.:s.n.

Lakshman, A. &. M. P., 2010. Cassandra: A Decentralized Structured Storage System', ACM SIGOPS Operating Systems Review. *ACM SIGOPS Operating Systems Review,* pp. 35-40.

Laney, D., 2001. *3D Data Management: Controlling Data Volume, Velocity, and Variety', Gartner Research.* [Online]
Available at: https://www.gartner.com/en/insights (A
[Accessed 2025].

Stonebraker, M. &. Ç. U., 2005. *One Size Fits All": An Idea Whose Time Has Come and Gone.* s.l., Proceedings of the 21st International Conference on Data Engineering, pp. 2–11..