ST10051335-TYRIQUE HANIFF INSY6112 ASSIGNMENT

Varsity College Durban North

Question 1

Recommended Database Type

For this social networking site, it is advisable to use a NoSQL database. NoSQL databases are ideal for dynamic and scalable web applications like social media since they are meant to store an enormous amount of unstructured and semi-structured data. (Han, Hai and Le, 2011).

Motivation for NoSQL Database

NoSQL databases support horizontal scalability, which means that instead of increasing the power of a single strong system, the system architecture allows for the expansion of user data and operational capacity by adding more servers (Sadalage and Fowler, 2012). This is crucial in maintaining fault tolerance and high availability (Han, Hai and Le, 2011).

NoSQL databases support a wide range of data structures like JSON, BSON, and key-value pairs compared to relational databases that conform to a fixed structural schema. This flexibility itself negates the need for schemas to be predefined, making it easier to store text data, images, videos, and user activity. (Moniruzzaman and Hossain, 2013).

NoSQL databases, specifically those designed to support quick read and write capabilities, allow real-time inspection of popular articles and interactions like comments, shares, and likes. This feature makes the user experience better and provides instant updates to user feeds. (Strauch, 2011).

Types of Data Stored in the NoSQL Database

Composed of data associated with the user profile, that is name, biography, photograph, areas of interest, past activities, etc. User-generated content such as wallpaper, images, videos, gifs, posts, items uploaded by users, comments. User Engagement Metric which are all the engagement statistics of your post, right from the likes to the shares and comments that your post receives from the users, will have to carefully be measured. This is a class of data that is supposed to use for the performance improvement and security monitoring. This becomes crucial in determining the different aspects in the system in which analytics decisions can be made in real time. (Sharma and Tim, 2016).

Types of NoSQL Databases

Document Repositories

Document databases, for instance, with MongoDB, provide flexible storage of data organized in JSON-like documents. These documents can have fields of varying data types, thereby becoming more suited for variable content provided by users such as comments and posts, as they do not require any design of the schema. This flexibility makes it very easy for operations like upgradation and modification. (Chodorow, 2013).

Key-Value Databases

Data storage technologies like Redis and DynamoDB represent an effective observation of key-value pair-based data model owing to their capability to cache data such as user session state, notifications, and real-time analytics etc. Data key-value stores have very low read latencies which help to provide a good experience for the user. (Gessert et al., 2017).

Column Family Databases

Examples would include Apache Cassandra and HBase, both column-store databases. These databases are particularly good with heavy analysis and time-series data stored as user activity logs and engagement metrics within different time ranges. (Lakshman and Malik, 2010).

Graph Database

Neo4j and Amazon Neptune serve as graph database engines designed specifically to model complex relationships between entities. They function well in storing social networks, friendship relations, and recommendation systems based on user interactions and relational patterns. (Robinson, Webber and Eifrem, 2015).

The Three Vs of Big Data in This Scenario

Volume

It generates huge amounts of data as a product of user actions, posts, likes, shares, comments such that an evergrowing amount of data may require efficacious handling and storage. This requires a one-stop solution for a No SQL database that will provide a perfectly scalable storage system that will meet the continuously growing data needs. (Zikopoulos et al., 2012).

Velocity

The architecture is designed for real time interaction without any hassle, where user feedback is important for an accurate push notification mechanism of trending content along with other alerts relevant to them. This critical requirement for real-time

interactions can be efficiently met by NoSQL databases with high read and write speeds. (Grolinger et al., 2013).

Variety

The various types of information are available in different formats such as (but not limited to) text documents, graphics files, audiovisual material, and organized metadata. It can thus be said that a NoSQL database creates essential flexibility, critical to proper handling and maintenance of these kinds of information. (Moniruzzaman and Hossain, 2013).

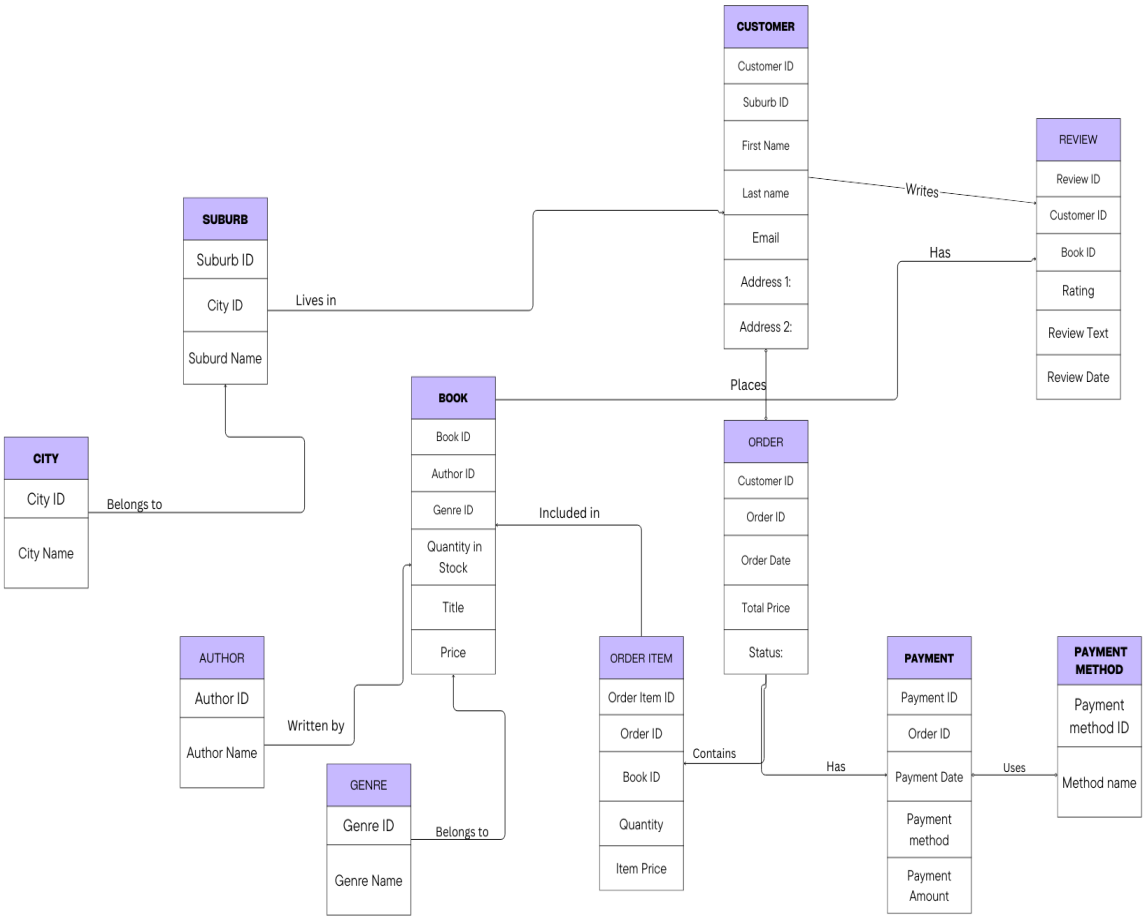
Alternative Solution: Relational (SQL) Databases

NoSQL databases are generally preferred, while a relational database of distributed architecture, e.g. PostgreSQL, or MySQL with clustering and sharding schemes is another equally good choice. This solution ensures structured data consistency and integrity while providing scalability through replication and partitioning mechanisms. (Stonebraker, 2010).

Conclusion

NoSQL databases are the ultimate solutions for large-scale social networking sites. This is due to the main features associated with NoSQL databases include high scalability, wonderful elasticity, and excellent analytical operation performance for real-time processing. Through the various classes of NoSQL databases available, social networking sites can successfully process and manage tons of user-generated information while also keeping track of user activities. It manages all the critical system data easily. All these are performed in a manner that makes sure users are provided with a seamless experience accessing the site.

Question 2



Bigger image on the last page after references for better perspective.

References

Question 1

Cattell, R. (2011). Scalable SQL and NoSQL Data Stores. ACM SIGMOD Record, 39(4), 12-27. https://doi.org/10.1145/1978915.1978919 (Accessed 26th March 2025)

Stonebraker, M. (2012). NewSQL: An Alternative to NoSQL and Old SQL for New OLTP Apps. Communications of the ACM, 55(5), 10-11. https://doi.org/10.1145/2160718.2160722 (Accessed 26th March 2025)

Han, J., Haihong, E., Le, G. and Du, J. (2011). Survey on NoSQL Database. In 2011 6th International Conference on Pervasive Computing and Applications (ICPCA), pp. 363-366. IEEE. https://doi.org/10.1109/ICPCA.2011.6106531 (Accessed 26th March 2025)

Sadalage, P. & Fowler, M. (2012). NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Addison-Wesley Professional. (Accessed 26th March 2025)

Chodorow, K., 2013. *MongoDB: The Definitive Guide*. 2nd ed. O'Reilly Media. (Accessed 26th March 2025)

Gessert, F., Wingerath, W., Friedrich, S. and Ritter, N., 2017. NoSQL database systems: a survey and decision guidance. *Computer Science - Research and Development*, 32(3-4), pp.353-365. (Accessed 26th March 2025)

Grolinger, K., Higashino, W.A., Tiwari, A. and Capretz, M.A.M., 2013. Data management in cloud environments: NoSQL and NewSQL data stores. *Journal of Cloud Computing: Advances, Systems and Applications*, 2(1), pp.1-24. (Accessed 26th March 2025)

Lakshman, A. and Malik, P., 2010. Cassandra: A decentralized structured storage system. *ACM SIGOPS Operating Systems Review*, 44(2), pp.35–40. (Accessed 26th March 2025)

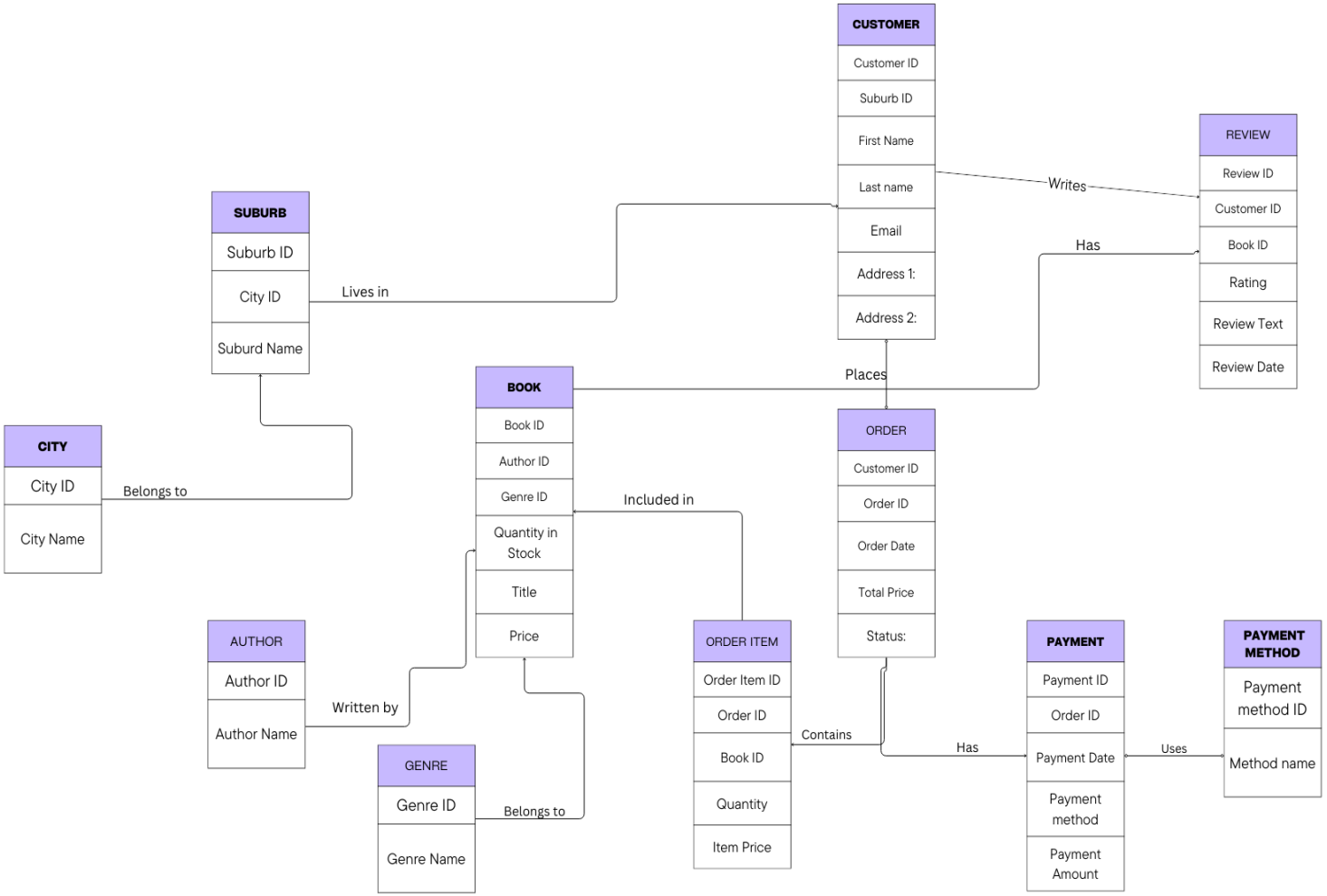
Moniruzzaman, A.B.M. and Hossain, S.A., 2013. NoSQL database: New era of databases for big data analytics - Classification, characteristics and comparison. *International Journal of Database Theory and Application*, 6(4), pp.1-14. (Accessed 26th March 2025)

Robinson, I., Webber, J. and Eifrem, E., 2015. *Graph Databases: New Opportunities for Connected Data*. 2nd ed. O'Reilly Media. (Accessed 26th March 2025)

Sharma, S. and Tim, U.S., 2016. Performance analysis of NoSQL and SQL databases for big data workloads. *International Journal of Computer Applications*, 143(3), pp.1-5. (Accessed 26th March 2025)

Strauch, C., 2011. NoSQL databases. *Lecture Notes, Stuttgart Media University*, 20(2), pp.1-52. (Accessed 26th March 2025)

Zikopoulos, P., Eaton, C., deRoos, D., Deutsch, T. and Lapis, G., 2012. *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw-Hill. (Accessed 26th March 2025)


CUSTOMER
Customer ID
Suburb ID
First Name
Last name
Email
Address 1:
Address 2:
REVIEW
Review ID
Customer ID
Book ID
Rating
Review Text
Review Date
Writes
Has
SUBURB
Suburb ID
City ID
Suburd Name
Lives in
Places
BOOK
Book ID
Author ID
Genre ID
Quantity in Stock
Title
Price
ORDER
Customer ID
Order ID
Order Date
Total Price
Status:
CITY
City ID
City Name
Belongs to
Included in
AUTHOR
Author ID
Author Name
Written by
GENRE
Genre ID
Genre Name
Belongs to
ORDER ITEM
Order Item ID
Order ID
Book ID
Quantity
Item Price
Contains
Has
PAYMENT
Payment ID
Order ID
Payment Date
Payment method
Payment Amount
Uses
PAYMENT METHOD
Payment method ID
Method name