PRACTICAL ASSIGNMENT 1

INSY6112 1B

MARCH 28, 2025

HIRANYA RAMLACKAN
ST10380113

Content Page

Question 1

To store, process and analyse the data from a large social media platform which gets used by millions of individuals around the world a NoSQL database would be ideal.

A NoSQL (Not Only SQL) database manages unstructured to semi-structured data that are in large amounts. This type of database supports horizontal scalability while providing flexible data models, this means that new servers can be easily added to accommodate for the increased data loads (GeeksforGeeks, 2025).

Unlike a relational database, a NoSQL database handles data coming in high velocity which is ideal for large social media platforms as there are millions of individuals around the world posting, commenting and interacting with the platform therefore bottlenecking is prevented.

In addition, this type of database operates on a global scale which means there are several locations and cloud centres used for reading from the database and writing to the database, therefore allowing for data to be easily distributed and always available. Notifications from social media platforms are needed to be fast, this is made possible through continuous monitoring which a NoSQL database provides.

Lastly, there is no specified scheme required to start using the application or platform, which means the entire database does not need to be redesigned when data is added. The NoSQL database allows for all the data types, which include text, images, videos, likes, and shares, to be stored together (Geeksforgeeks, 2023). Therefore, NoSQL databases are cost-effective, easy to design, functional and perform well when used at scale (Amazon).

There is a variety of different types of data that can be stored in a NoSQL database such as documents, key-value, wide columns, in-memory, and graphs.

Document databases store semi-structured data in a JSON-like document which also gets queried. This makes creating and updating the platform easier.

Key-value databases assign a unique key to every value. These values consist of string, number, Boolean, or complex objects. The key is used to store and retrieve the values.

Wide column databases store and read data in rows, which are organised in columns and each columns name and format varies. This is ideal when a specific value from a given column can be located easily and quickly.

In-memory databases store the data in memory. This eliminates lagging in real-time applications (Google Cloud).

There are 4 recommended databases for this particular scenario, these include time-series databases, graph databases, relational databases, and cloud databases.

Time-series databases store data that has been monitored over time (influx data).

Graph databases focus on the relationship between the various data in the database. This data is organised based on the relationships it has with other data in the database. This type of database has enhanced performance and adaptability (Amazon).

Relational databases store data in rows and columns forming multiple tables. Each table consists of data that is related to each other and has a primary and foreign key (IBM, 2021).

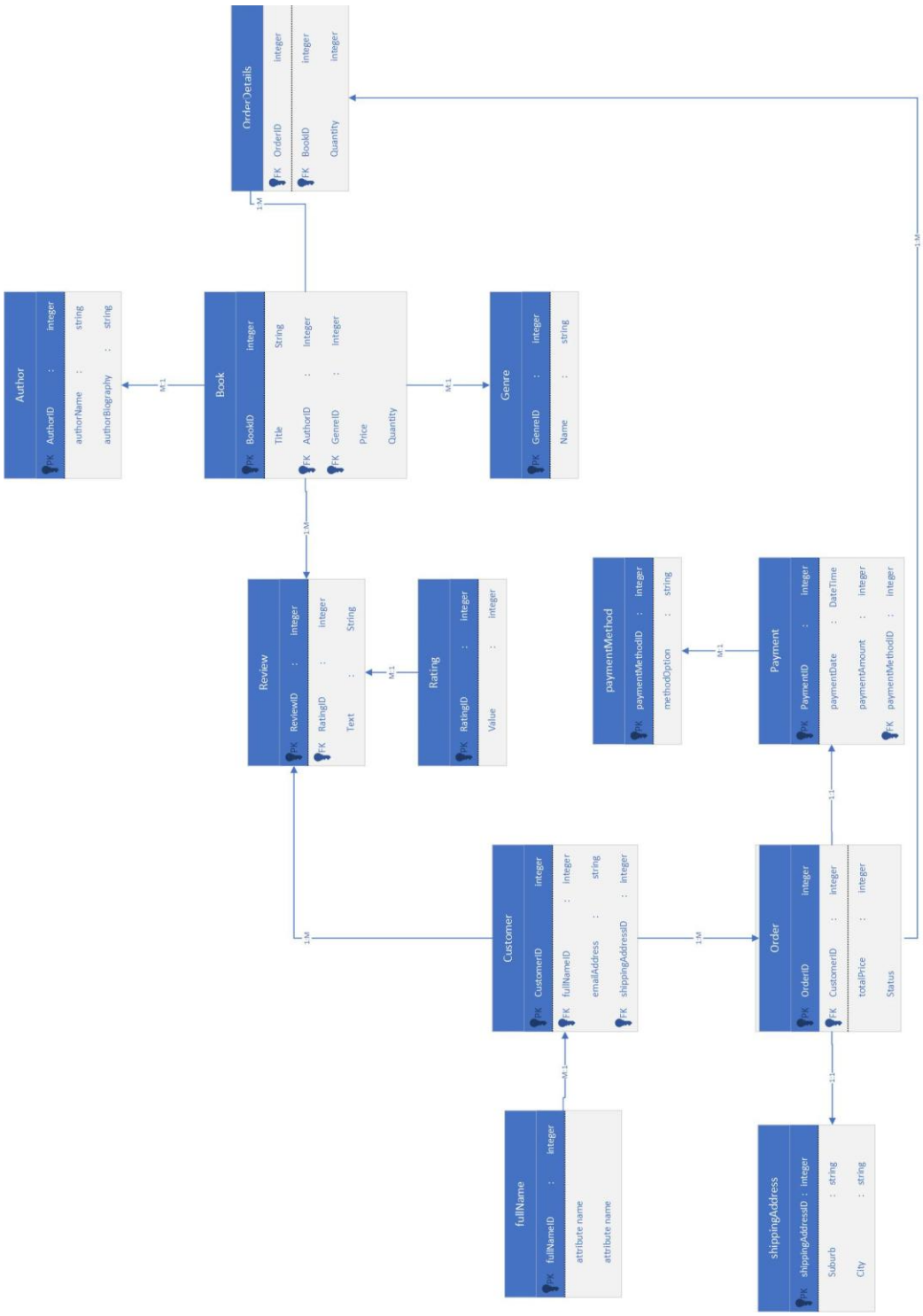
Cloud databases use a cloud computing platform to build and access the database from. This type of database is flexible as the required software is installed on a cloud infrastructure (IBM).

The three V's of big data are volume, velocity and variety. Volume refers to the amount of data. Social media platforms collect millions of data daily. This consists of posts, likes, shares, images, videos, and comments. As more people join the platform, more data is collected, therefore the volume increases and suitable storage that will hold this amount of data is required.

Thereafter, velocity of the data is taken into consideration. Velocity refers to the speed at which the data is processed and generated. Since social media platforms create, update and access data instantly, this type of platform processes data in real time. This allows users to receive live updates from the platform.

Lastly, variety, this refers to the type of data, whether it is structured, semi-structured or unstructured. Data used on social media platforms are text, multi-media and user interactions. Structured data is data that has a defined length and format and is organised, semi-structured data does not comply with the data structure and is only semi-organised, and unstructured data is data that does not neatly fit into rows and columns and is unorganised (GeeksforGeeks, 2023).

Question 2


OrderDetails
OrderID integer
BookID integer
Quantity integer
Author
AuthorID integer
authorName string
authorBiography string
Book
BookID integer
Title String
AuthorID Integer
GenreID integer
Price
Quantity
Genre
GenreID integer
Name string
Review
ReviewID integer
RatingID integer
Text String
Rating
RatingID integer
Value integer
paymentMethod
paymentMethodID integer
methodOption string
Payment
PaymentID integer
paymentDate DateTime
paymentAmount integer
paymentMethodID integer
Customer
CustomerID integer
fullNameID integer
emailAddress string
shippingAddressID integer
Order
OrderID integer
CustomerID integer
totalPrice integer
Status
fullName
fullNameID integer
attribute name
attribute name
shippingAddress
shippingAddressID integer
Suburb string
City string

Bibliography

Introduction to nosql (2025) *GeeksforGeeks*. Available at: https://www.geeksforgeeks.org/introduction-to-nosql/ (Accessed: 17 March 2025).

Top 5 reasons to choose nosql (2023) *GeeksforGeeks*. Available at: https://www.geeksforgeeks.org/top-5-reasons-to-choose-nosql/ (Accessed: 17 March 2025).

What is a nosql database? - nonrelational databases explained - AWS (no date) *aws*. Available at: https://aws.amazon.com/nosql/ (Accessed: 26 March 2025).

What is nosql? databases explained | google cloud (no date) *Google*. Available at: https://cloud.google.com/discover/what-is-nosql (Accessed: 18 March 2025).

Time Series Database (TSDB) guide: Influxdb (2025) *InfluxData*. Available at: https://www.influxdata.com/time-series-database/#what-is (Accessed: 25 March 2025).

What is a graph database? - graph DB explained - AWS (no date) *aws*. Available at: https://aws.amazon.com/nosql/graph/ (Accessed: 25 March 2025).

What is a relational database? (2021) *IBM*. Available at: https://www.ibm.com/think/topics/relational-databases (Accessed: 26 March 2025).

What is a cloud database? (2025) *IBM*. Available at: https://www.ibm.com/think/topics/cloud-database (Accessed: 26 March 2025).

6V's of Big Data (2023) *GeeksforGeeks*. Available at: https://www.geeksforgeeks.org/5-vs-of-big-data/ (Accessed: 26 March 2025).