

Prashik Gangaparsad
ST10401917

INSY6112
Practical Assignment

IIE Varsity College Pietermaritzburg

Content Page

Page 1 - Content Page

Page 2 – Motivation & Kinds of data stored in the database

Page 3 – Types of databases and the three Vs of big data

Page 4 – Entity Relationship Diagram

Page 5 - Small description of the entity relationship diagram

Page 6 – Reference List

Report

Based on the platforms requirements, not only SQL also called (NoSQL) database should be used. NoSQL follows an untraditional approach when it comes to storing data. Unlike SQL in traditional database that uses tables to store data, NoSQL does not use tables to stores and query data (IBM, 2022).

Motivation

Comparing to relational databases, data is stored in tables before the data can be access and analysed by the system. Relational databases are usually strict on how the data is structured and that means that the data must be arranged in a proper manner before it can be stored (Spiceworks, 2022). NoSQL databases does not have any of these limitations. They can handle different types of data structures which include unstructured, semi-structured, and structured data without having any issues. This makes NoSQL very flexible especially for applications that needs different data models (Spiceworks, 2022). Since NoSQL supports different data models, the same data can be used in multiple ways without needing separate databases which makes it easier and faster to manage data across the network (Spiceworks, 2022).

Comparing to relational databases, NoSQL databases can copy data almost instantly across all their different nodes. Each node can have several copies of itself which can take over if the primary node fails. The nodes including their copies can be located in different parts around the world which helps make sure that the system always stay online even if a server crashes, a power outage occurs, or even a natural disaster occurs (Macrometa, 2025).

Modern day NoSQL databases can now distribute data on a global scale. This is achieved by using multiple cloud regions and data centers which allows the users to read and write data from different locations (Spiceworks, 2022). On the other hand, because relational databases usually will use and depend on a single location for storing and accessing data, this causes much more slower operations than usual for the users that are far away from the database (Spiceworks, 2022). NoSQL databases solve this issue by making copies of the data and storing them in different locations which makes the information always available to where it is needed thus reducing wait times and making the access of data much faster (Spiceworks, 2022). NoSQL uses a method called sharding which is used to spread the data across all different machines. As this will help to keep the data organised and makes it easier to handle much larger amounts of information (Spiceworks, 2022).

Kinds of data

NoSQL database will store different kinds of data for the social media platform, each user in the system will have a unique identifier which will act as the primary key. User will also have a username that will be unique to them and an email address will be used for authentication (GeeksforGeeks, 2024). To keep their account safe and secure, users will have a password. The system will also store the user's full name and an optional bio (description) about themselves. Also users can have a profile picture which is saved as a URL or a reference in the system (GeeksforGeeks, 2024).

Each post in the system will have a unique identifier which will act as the primary key which will be linked to the user's primary key as it will identifies the user who created the post (GeeksforGeeks, 2024). The text of the post will be stored as well. If the post includes media, whether it is an image or video, the URL or a reference will store the link to the media file and a timestamp to record the exact date and time when the post was created (GeeksforGeeks, 2024).

Each comment in the system will have a unique identifier which serves as the primary key which will be linked to a post's primary key which will show which post the comment belongs to and the user's primary key which will identify the user who wrote the comment (GeeksforGeeks, 2024). The comment itself will be stored in a field which will contain the text of the comment and a timestamp to record the exact date and time when the comment was posted (GeeksforGeeks, 2024).

Each like in the system will have a unique identifier which serves as the primary key. which will be linked to a post's primary key to show which post was liked (GeeksforGeeks, 2024). The user's primary key identifies the user who gave the like and a timestamp to record the exact date and time when the like was added (GeeksforGeeks, 2024).

Types of NoSQL database

There are four types of NoSQL databases: Document Store, Key-value, column store and graph store databases. A document store database is basically a document that stores records in a document-based database. The database will contain information about a type of data along with any extra information which is called metadata (Tutorialspoint, 2023). Each document is made up of field value pairs which can store different types of data such as arrays, strings or integers. These documents are usually saved in formats such as XML, JSON, or BSON (Tutorialspoint, 2023). Every document has a unique ID which makes it easier to find and process data in the document store database. One of the main advantages of document store databases is that the database does not require a fixed structure. This means that documents in the same collection can have different fields. Because different types of data can be stored as a collective, this makes the document store database very flexible. (Tutorialspoint, 2023).

Key-value databases stores data in a simple key value format. Each key is unique and is linked to a specific value. The value can be anything like for example a string or a number (Airbyte, 2024). These databases are quite fast because users can quickly access any data using the key without needing any complex queries which makes them great for handling large amounts of data as the database can quickly read and write operations (Airbyte, 2024).

A column store database is a type of database that stores data in columns instead of rows. In a traditional database where the data is saved in rows but in a columnar database each column is stored separately as this makes it faster and more efficient when searching for data in a specific column because the database does not have to go through the entire row to find what is needed (Airbyte, 2023). Because columns can be stored separately, this gives flexibility in adding or modifying the columns without interfering with any other columns. (Airbyte, 2023).

A graph store database is designed to store data and its relationships in a way that is easy to understand and use. Instead of storing data in tables like in a traditional database, a graph database organises the data as nodes, which are entities and edges, which are relationship between them as this makes it easier to understand how data are connected in the real world (NebulaGraph, 2023). Graph databases are great for working with complex data because they quickly find relationships between different pieces of information as that rely on understanding connections between data. This structure makes it easy to store and retrieve data that has complex connections in the system (NebulaGraph, 2023).

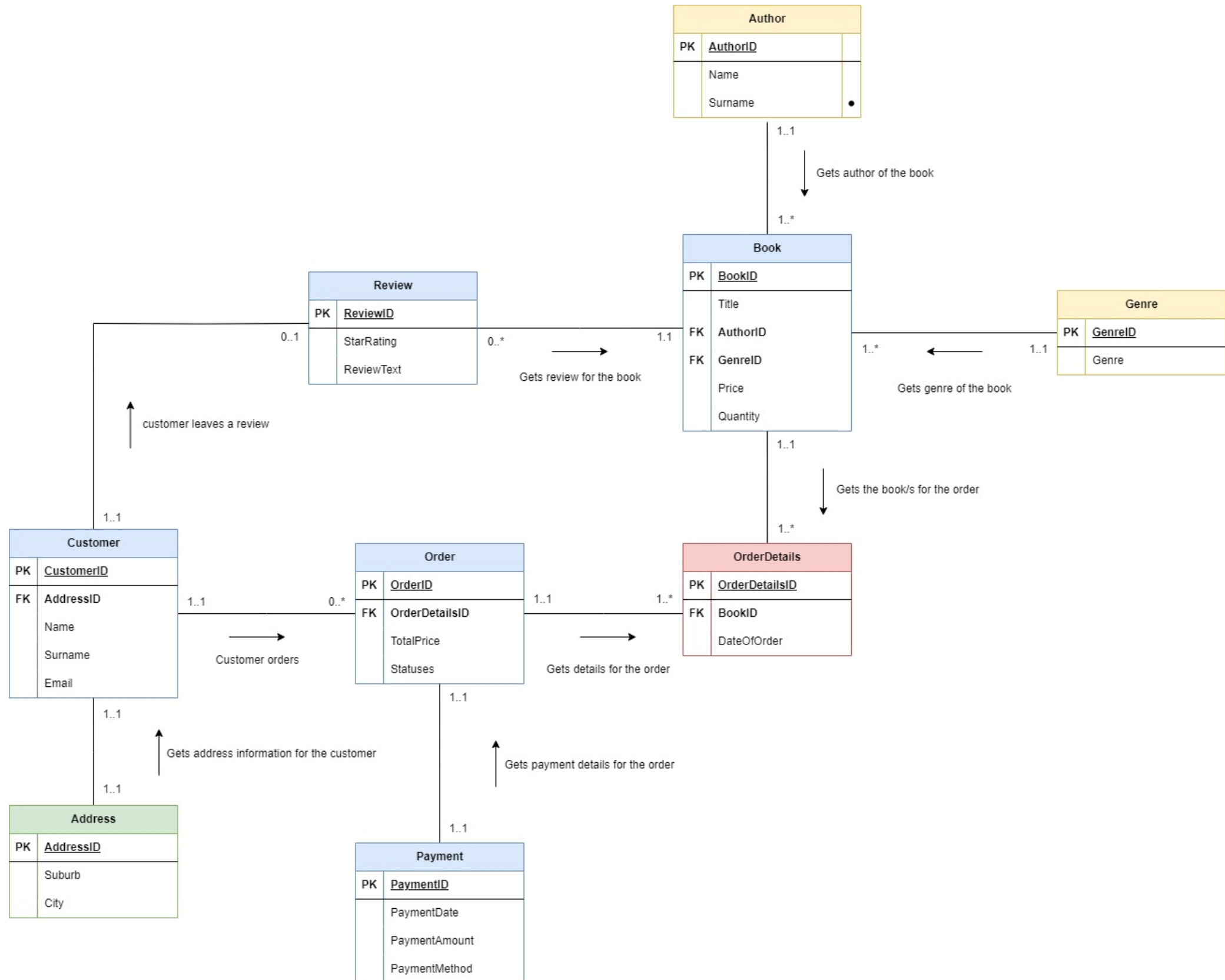
Three Vs of big data

There are three Vs of big data is called: volume, velocity and variety. Volume refers to how big the data is (Google, 2025). This is going to be a large social media platform so the volume of data will be enormous as thousands of users will create accounts, content and interact with it on a daily basis in the form for posts, comments, likes, videos, live streaming and photos which will all generate big data (Google, 2025).

Velocity refers to how fast the data has been generated (Google, 2025). Velocity is very important in a social media platform because the data generated on the platform will be a continuous process. The trending topics, hot posts and engagement metrics will need to be updated almost instantly. Users can also expect real-time interaction on their feeds, including notifications, likes, and comments (Google, 2025).

Variety refers to how the data has been structured (Google, 2025). The social media platform will have a variety of different data structures because there will be types of contents and the data associated with these interactions can come in various forms. Each type of content requires different data structures to be stored and processed (Google, 2025).

Entity Relationship Diagram (ERD) (NEXT PAGE)



Small description of the entity relationship diagram

- Customer entity is associated with three other entities: Order, Address, Ratings.
 - The address entity contains the suburb and city.
 - The foreign key AddressID in the customer entity will contain the suburb and city.
-

- The order entity is associated with OrderDetails and Payment.
 - Since order and book had a many to many relationships, a bridging entity called OrderDetails was created to split the relationship up.
 - OrderDetails contains the BookID and the date of the order.
-

- The Book entity has two foreign keys, the AuthorID and GenreID.
 - The reason why an author entity was created is that an author can write many books but a book can be only written by one author.
 - Same reasoning with genre. There are many genres a book can be but only one genre can be associated with one book only.
-

- The review entity has a star rating and a review text for a small description. A customer can only write only one per book but book can have many reviews. A customer can write many reviews.

Reference List

- Airbyte, 2024. *Types of NoSQL Databases: Unleashing Data's Full Potential*. [online] Available at: <<https://airbyte.com/top-etl-tools-for-sources/types-of-nosql-databases>> [Accessed 28 March 2025].
- Airbyte, 2023. *Columnar Database Storage Formats: Benefits and Use Cases*. [online] Available at: <<https://airbyte.com/data-engineering-resources/columnar-storage>> [Accessed 28 March 2025]
- GeeksforGeeks, 2024. *How to Design Database for Social Media Platform*. [online] Available at: <<https://www.geeksforgeeks.org/how-to-design-database-for-social-media-platform/>> [Accessed 28 March 2025].
- Google, 2025. *What is Big Data?*. [online] Available at: <<https://cloud.google.com/learn/what-is-big-data>> [Accessed 28 March 2025].
- IBM, 2022. *What is a NoSQL database?*. [online] Available at: <<https://www.ibm.com/think/topics/nosql-databases>> [Accessed 28 March 2025].
- Macrometa, 2025. *Advantages and disadvantages of NoSQL*. [online] Available at: <<https://www.macrometa.com/distributed-data/advantages-and-disadvantages-of-nosql>> [Accessed 28 March 2025].
- NebulaGraph, 2023. *What is a Graph Database and What are the Benefits of Graph Databases*. [online] Available at: <<https://www.nebula-graph.io/posts/why-use-graph-databases>> [Accessed 28 March 2025].
- Spiceworks, 2022. *What Is NoSQL? Features, Types, and Examples*. [online] Available at: <<https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-nosql/>> [Accessed 28 March 25].
- Tutorialspoint, 2023. *Document Database in NOSQL*. [online] Available at: <<https://www.tutorialspoint.com/document-database-in-nosql>> [Accessed 28 March 2025].