**INSY6112 A1**

**Nikhil Ramesar**

**ST10482284**

**Question 1**

A NoSQL database is just a database of the non-relational type enabling both the storage and also retrieval of data that is unstructured, semi-structured, and structured. Relational databases use fixed schemas within tables; however, NoSQL databases are flexible. They support data under multiple formats including documents key-value pairs wide columns and graphs. In applications, scalability, fast access, and dynamic handling of complex data are requirements.

**Motivation for using NoSQL**

1.⁠ ⁠Scalability

NoSQL databases support horizontal scaling meaning adding servers is easy as data and user volume grows. This is critical for any social media platform in which millions of people newly interact every second.

2. Data Structures that happen to be Complex Benefit because of Flexibility

Users generate content such as text images videos and even live streams. NoSQL databases store all of this data without the need for enforcing some rigid schema which does easily make the adding of new content types and features possible over some period of time.

3. Real-Time acts as a category. Performance does belong within this category.

NoSQL systems such as Cassandra and MongoDB are optimized for quick read operations plus quick write operations since they support real-time analytics. This improves user experience thereby ensuring instant updates to trending topics likes notifications and feeds.

**Types of Data Stored in the Database**

User data is made of profiles, account settings, preferences, and follower connections. User profiles plus account settings comprise this data.

• Posts, also photos, plus videos, and even live streams, with stories: Content Data.

• Likes, shares, comments, tags, and Reactions: Interaction Data.

• Activity logs, engagement statistics, as well as trending topics: Analytics Data.

**Four Types of NoSQL Databases**

1. Databases of Documents

Databases of documents use documents that are JSON-like for storing information. These documents include key-value pairs arrays and nested objects. Each document can represent either a post, a comment, or an user profile. They can be ideal for some social media content because they do things such as that. They provide flexible schemas which allow adding new attributes without disrupting the existing data structure. MongoDB popularity exemplifies this database type.

2. Key-Value Stores

A key-value store saves data since pairs of keys and values exist because each key maps directly to a value. They are fast as well as efficient for more simple lookups. User sessions, cache feeds, or login tokens can be stored by them. Well-known examples DynamoDB and Redis support high-performance applications.

3. Column-Family Stores

Column-family databases handle large amounts of time-series and analytical data efficiently because of data organization into columns rather than rows. They are in fact useful ones for the storing of metrics that are like likes per minute. They are useful also in the storing of trending hashtags as well as user engagement patterns. A database that is widely used is Apache Cassandra. It is a column-family database too.

4. Graph Databases

Graph databases stress relationships between entities such as users and connections. They represent data as nodes such as users or posts. They represent data also as edges like friendships, likes, shares. Interactions and also relationships are central in that this structure is ideal for all social networks. For such of the purposes, Neo4j is a well-known graph database.

**The 3 Vs of Big Data in the Scenario**

1.⁠ ⁠Volume

Worldwide millions of users generate billions of posts, comments, and interactions daily. This platform is always processing the large data amount.

2.⁠ ⁠Velocity

Very fast pace makes and revises data. Feeds, notifications, and trending analytics do require real-time updates, which means data must be processed in an almost instant manner.

3.⁠ ⁠Variety

Data exists as text, images, videos, live streams, and user reactions. For structured, semi-structured, also unstructured data, a flexible system is required for the handling of this diversity.