Quiz 2 – A
Duration: 30 minutes

Q1) Please write down your name and student ID [10 points].

Q2) Choose the correct answer [30 points].

- It is a smart broker with simple consumers, and it supports multiple types of exchange.
    - o ==RabbitMQ==
    - o Kafka

- This component makes Spark works faster and more efficient than Hadoop.
    - o ==Resilient Distributed Datasets (RDD)==
    - o A distributed file system

- What are the properties of the CAP theorem?
    - o Consistency, Adaptability, and Partition Tolerance
    - o Compatibility, Availability, and Performance
    - o ==Consistency, Availability, and Partition Tolerance==

- Point out the **WRONG** statement about MapReduce.
    - o MapReduce is a programming model for the distributed analysis of big data
    - o MapReduce follows the "primary-secondary" architecture
    - o ==MapReduce works well for real-time streaming applications==

- Point out the **WRONG** statement about Lambda and Kappa architecture.
    - o ==In Kappa, we reprocess each data in streaming and batch layer==
    - o In Lambda, the code is replicated to handle the same tasks but for different types of data

Please check the other side for the rest.

Q3) Write down the name of your favorite book (I know, so many 😊, choose one). The following answer are also acceptable: prefer not to share, I don't have any, hmmm so hard to choose, etc. [10 points]. Hmmm, so many.

Hmm, so many. Anne of Green Gables Series, to kill a mocking bird, Isabel Allende's books, and many more.

Q4) You develop a system for holding patient records. Most records concern files (patient history, drug prescriptions, exams). The files can come from different hospitals, clinics, or pharmacies so their contents may differ. For analysis purposes, we may want to be able to search files by their contents (for example find all patients with high blood pressure). We already have these data stored in relational databases. What kind of NoSQL database would you use to implement the system? Justify your response. (50 points)

The answer is Graph Databases.

- Complex Relationships: Social networks involve intricate relationships between users. Users are connected to each other through friendships, follow relationships, group memberships, and more. Graph databases are designed to efficiently model, store, and traverse such complex relationships.
- Recommendation Algorithms: Graph databases support graph-based algorithms, such as collaborative filtering and personalized PageRank, which are essential for generating meaningful friend suggestions based on mutual connections and user behavior.
- Real-Time Recommendations: Graph databases excel at providing real-time recommendations as users interact with the platform. They can quickly identify and suggest new connections as users add friends, join groups, or interact with content.
- Scalability: As social networks grow; the graph database can scale horizontally to accommodate increasing numbers of users and relationships while maintaining fast query performance.
- Query Efficiency: Graph databases are optimized for graph traversal queries, making it efficient to find connections and discover potential friends for users.

By choosing a Graph Database like Neo4j or Amazon Neptune, you can build a recommendation system that harnesses the power of graph theory to deliver relevant and timely friend suggestions to users, enhancing their social media experience.