Department of Computer Science and Engineering

Faculty of Engineering

University of North Texas

Assignment 3 CSCE5300 Spring 2024

Due on or before 5th April 2024.

**Objective**: To provide students with hands-on experience in designing a Cassandra database schema, performing data modeling, and optimizing queries for efficient data retrieval. This assignment involves self-learning component.

Tasks:

1. Database Design and Schema Creation (20 points):
   * Use case: social media application.
   * Design a Cassandra database schema to accommodate the data requirements of the chosen use case scenario.
   * Define appropriate partition keys, clustering keys, and secondary indexes based on the anticipated query patterns.
   * Justify the design decisions made and explain how they align with the requirements of the use case scenario.
2. Data Modeling and Population (30 points):
   * Populate the Cassandra database with sample data.
   * Generate synthetic data or use publicly available datasets if applicable.
   * Implement data modeling techniques such as denormalization, materialized views, and collections to optimize data storage and retrieval.
   * Include a brief description of the data population process and any challenges encountered.
3. Query Optimization and Performance Tuning (40 points):
   * Formulate a set of representative queries that cover a variety of query patterns (e.g., read-heavy vs. write-heavy queries, range queries, filtering by partition key or clustering key).
   * Execute each query against the populated Cassandra database and measure the query performance.
   * Identify any inefficient queries and propose optimization techniques to improve their performance.
   * Implement query optimizations such as denormalization, indexing, caching, and query restructuring.
   * Compare the performance of optimized queries against the original queries and present the results.
4. Documentation and Presentation (10 points):
   * Prepare a comprehensive report documenting the entire database design process, data modeling decisions, query optimization strategies, and performance evaluation results.
   * Include detailed explanations, code snippets, query execution plans, and performance metrics.

Submission Guidelines:

* Submit the assignment report in PDF format via canvas.
* Include references to any external sources or literature consulted during the assignment.

Evaluation Criteria:

* Adherence to assignment instructions and guidelines.
* Clarity and depth of database design and schema creation.
* Effectiveness of data modeling techniques and data population process.
* Creativity and effectiveness of query optimization strategies.
* Quality of documentation.