Name: **Nurlan Imanov (G49119771)** **August 6, 2023**

Title: **Weekly Report 5(Designing database to efficiently store and retrieve massive amounts of data)**

1 Introduction

This week, I set up the server that will conduct tests related to our project, which focuses on handling Big Data. My current efforts are aimed at evaluating various databases under Big Data scenarios, beginning with RDBMS. To accurately assess the situation, I am artificially creating different quantities of records to simulate the volume aspect of Big Data. The performance is then evaluated by measuring the execution time of complex queries under varying conditions. Details of these specific scenarios, along with insights into the results, will be elaborated in the sections below.

2 Initial Assessment and Selection of RDBMS

2.1 Real Data Selection

For the initial tests, 800,000 real datasets from the company's data were selected. This dataset effectively symbolizes the current data structure and serves as a reliable foundation for experimentation.

2.2 Performance Testing Methodology

Various queries were executed on the selected dataset to measure performance metrics, focusing specifically on execution time. The approach adopted from the research paper "Handling Big Data in RDBMS" guided the testing methods.

2.3 Optimization Techniques

Three distinct optimization techniques were explored:
1. **Table Partitioning**: Dividing tables into more manageable parts.
2. **Indexing**: Creating indexes to improve data retrieval.
3. **Combination of Table Partitioning and Indexing**: A hybrid approach to leverage both techniques.

2.4 Initial Results with 800,000 Records

Currently, I have tested only the Simple Approach where mo optimization yielded. I will add the metrics as updated version of report 5(Since I am currently working on it)

3 Planned Steps for Continued Testing

The project's next phases include replicating the tests for 1,600,000; 4,000,000; and 8,000,000 records to assess the scalability and further refine our understanding of RDBMS's performance limitations.

4   Future Work on NoSQL Testing with MongoDB

The subsequent stage of this investigation will involve conducting similar tests on one of the NoSQL databases, specifically MongoDB. This phase is vital to drawing accurate comparisons between RDBMS and NoSQL, helping to identify the best-suited database for our Big Data requirements.

5   Conclusion

The progress made thus far forms the groundwork for a comprehensive analysis of the company's existing database system. By methodically evaluating RDBMS and implementing optimization techniques, the company can be ready to look into moving to NoSQL solutions if necessary. The insights gained from this process will be instrumental in steering the company's strategic decisions regarding database design, aligning with future growth and Big Data challenges.