



Name: **Nurlan Imanov (G49119771)**

June 4, 2023

Title: **Designing database to efficiently store and retrieve massive amounts of data**

1 What are you going to do?

I will undertake a comprehensive study of database design focusing on large-scale applications. The study will investigate the efficient storage and retrieval of massive volumes of data, with special attention to textual data. The research will cover both relational and NoSQL databases. I will investigate what are the current approaches what are the limitations pros and cons of existing solutions.

2 How is it done today? Current Limitations?

Currently, different database systems are utilized to handle massive data. For relational databases, techniques such as partitioning, indexing, and data normalization are used. In NoSQL databases, strategies like sharding, replication, and denormalization are prevalent. Distributed databases make use of horizontal scaling and data replication to manage large data sets. Despite these techniques, they come with their limitations including difficulties in managing complex queries, consistency issues, difficulties in achieving ACID compliance, and managing trade-offs as per CAP theorem.

3 What is your idea to do something better?

My plan is to deeply study the existing methods of handling large amounts of data in databases, identify their strengths and weaknesses, and propose an optimal combination of strategies that can deliver better performance in terms of data storage, retrieval, fault tolerance, and IO optimization. This could involve novel database schema designs, innovative use of data structures and algorithms, or hybrid approaches combining the best aspects of both relational and NoSQL databases.

4 Who will benefit from your work? Why?

Database administrators and system architects can use the insights to design more efficient databases. Developers can leverage the strategies proposed to optimize their applications. Organizations can benefit from the efficient, reliable, and scalable database designs that will ensure high performance even with massive data volumes, ultimately improving customer satisfaction.

5 What risks do you anticipate?

The risks associated with this research include the challenge of simulating real-world conditions for testing.

6 Out of pocket costs? Complete within 11 weeks?

The theoretical part of the research should be feasible within the 11 weeks timeframe. I can investigate what are the current methods their limitations main obstacles. Novel approaches and experiments can be done throughout the master thesis

7 Midterm results?

By the midterm point, I aim to have a substantial review of existing methods and preliminary findings on their advantages and drawbacks. At this point, I should also be able to sketch out some initial directions for the proposed optimal database design.

8 Final Demonstration?

The final demonstration will present the results of the analysis, including a comparison of the existing methods.