IMPROVING THE PERFORMANCE OF REAL-TIME DATA ANALYTICS APPLICATIONS BY OPTIMISING THE DATABASE AGGREGATION

Thennamurage Dona Methma Pabasarie Samaranayake

199359L

MSc in Computer Science

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

June 2022

IMPROVING THE PERFORMANCE OF REAL-TIME DATA ANALYTICS APPLICATIONS BY OPTIMISING THE DATABASE AGGREGATIONS

Thennamurage Dona Methma Pabasarie Samaranayake

199359L

This thesis was submitted in partial fulfilment of the requirements for the Degree of MSc in Computer Science Specialising in Software Architecture

MSc in Computer Science

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka
June 2022

DECLARATION

I declare that this is my own work and this has not incorporated without acknowledgement any material previously submitted for Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. Also, I hereby grant to the University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other media. I retain the right to use this content in whole or partially in future works.

....

T. D. M. P. Samaranayake Date

I certify that the declaration above,
The candidate is true to the best of my knowledge and that this project report is acceptable for evaluation for the final research.

....

Prof. Indika Perera Date

ABSTRACT

Organisations must make the best decision at the appropriate time to obtain a competitive advantage in a fast-changing market. To accomplish so, it's critical to make faster and more efficient judgments based on near-real-time data analysis. When it comes to these real-time streaming data analysis systems, the performance of the database is having a huge impact on such applications as it is required to achieve data availability and continuous processing for a large volume of data without a delay. When it comes to streaming data, data warehousing is more challenging. So, it is required to consider performance improvements in all the steps of the Extraction, Transformation, Load (ETL) process and the database architecture level. Therefore, the proposed approach is to improve the performance of the system by optimising the ETL process (Extraction, Transformation, and Load) and real-time data warehousing. In this approach, the optimised aggregation algorithm is introduced. Apart from that, the hardware, storage schemas, and query optimization of the data warehouse are also considered and this study is evaluating the performance of the centralised architecture for the real-time data warehouse.

ACKNOWLEDGEMENT

My profound gratitude goes to Prof. Indika Perera, my supervisor for the knowledge, supervision, advice, and guidance provided with his expertise, throughout in making the thesis a success. My appreciation goes to my family for the motivation and support provided throughout my life. I also would like to thank my colleagues in the MSc batch and at my workplace for their help and support in managing my research work.

THE TABLE OF CONTENT

LIST OF FIGURES