DRDP : DYNAMICALLY RE-CONFIGURABLE DATA PIPELINE IN THE EDGE NETWORK

M.G.I.M. Nuwanthilaka

219376N

Master of Science in Computer Science

Department of Computer Science Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

July

DRDP : DYNAMICALLY RE-CONFIGURABLE DATA PIPELINE IN THE EDGE NETWORK

M.G.I.M. Nuwanthilaka

219376N

Thesis submitted in partial fulfillment of the requirements for the degree Master of Science in Computer Science

Department of Computer Science Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

July

DECLARATION

I declare that this is my own work and this Thesis does not incorporate without acknowledgement any material previously submitted for a 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. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date: 2023/07/1

The supervisor should certify the Thesis with the following declaration.

The above candidate has carried out research for the Master of Science in Computer Science Thesis under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Prof. Indika Perera

Signature of the Supervisor: Date: 13/07/2023

DEDICATION

Dedicated to all my past and present teachers.

ACKNOWLEDGEMENT

I would like to extend my sincere gratitude towards my supervisors Prof. Indika Perera and Dr. Gayashan Amarasinghe at the Department of Computer Science and Engineer- ing, University of Moratuwa, for their patience, dedication and guidance through out the research. Without their guidance, supervision and the expertise, this research could not have been completed.

Also I extend my gratitude to all the staff from the Department of Computer Science and Engineering who help me to improve my knowledge in various subject areas in multiple ways during this study.

I would like to express my heartfelt appreciation to my family members and friends in my study group. I am grateful to them for their understanding and encouragement throughout the course to make this research a success.

I wish to express my gratitude to all my colleagues at MillenniumIT ESP Pvt. Ltd and Cut+Dry,Inc for the support given to manage my MSc research work.

ABSTRACT

Pipelines are a highly discussed topic in today's technological world. There are different variations of pipelines; Data Science pipelines, DevOps pipelines, and DevSecOps pipelines, etc. A data science pipeline usually comes with a fixed architecture, which can be problematic in a fast-growing tech industry.Traditional data science pipelines may struggle to handle the volume, velocity, and variety of data at the edge, necessitating more dynamic and adaptable approaches. Many advancements are happening to bring the technology to the edge due to substantial data points generated in the sensor networks at the edge; from factory floors to log streams.

So, in this thesis we first discuss the existing literature in the data pipeline domain under three main topics; data pipeline challenges, data pipeline architectures, and data pipeline security. Then we propose a methodology for dynamically re-configurable data pipeline architecture in the edge network. This way we expect to achieve more efficiency, controllability, and scalability of the data across networks.The emerging field of edge architecture presents opportunities for innovative approaches to data pipelines, enabling organizations to harness the full potential of edge data for advanced analytics, machine learning, and real-time decision-making.Further, we propose a prototype with Raspberry Pi-based programs to discuss the effectiveness of this novel method. Using this proposed architecture we have evaluated the results and later discussed how this benefits the current and future data pipeline implementation. We hope this contributes to the emerging edge architecture subject area.

Keywords: data science, pipeline, architecture, edge

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES