Container Scheduling Optimization in a Clustering Environment

H G P Madhushankha

(219360K)

Degree of Master of Science in Computer Science

Department of Computer Science
Faculty of Engineering

University of Moratuwa
Sri Lanka

June 2024

DECLARATION

I declare that this is my own work, and this thesis/dissertation does not incorporate without acknowledgment 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 acknowledgment is made in the text.

Also, I hereby grant to the University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date: 16/06/2024

The above candidate has carried out research for the Master's dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of the Supervisor: Prof. Indika Perera

Signature of the Supervisor: Date: 17/06/2024

ACKNOWLEDGEMENT

This research has been made with the intention of deriving a solution in the microservice world which is improving day by day with the efforts of different fields of expertise. For the enhancement of this research, there were several people who spent their time and helped with their capabilities. I would like to offer my heartfelt gratitude to all of them including my supervisor Prof. Indika Perera, Department of Computer Science & Engineering, University of Moratuwa. And special thanks to my parents and my colleagues who encouraged me all the time.

ABSTRACT

The progress in container orchestration technology has led to a saturation level of microservices related solutions in recent years. Web applications typically exhibit a fluctuating workload that is challenging to estimate. Based on system parameters, the microservice design has the capability to dynamically alter the number of containers, either manually or automatically. There is a persistent issue regarding the adjustment of the number of containers in real time to match the current workload of the microservice. Several strategies have been utilized to tackle this difficulty effectively and swiftly, particularly in situations where there is a sudden and significant increase or decrease in the demand for services.

Scheduling the container resources according to the demand is one of the most focused requirement of the industry. Service providers and researchers are attempting to decrease the expenses while upholding the Quality of Service. Among the autoscaling approaches rule-based approach, machine learning models, Fuzzy rules, control theory and hybrid approaches have been discussed in previous studies. To deal with the above research problem, a container scheduling strategy as a hybrid approach, based on machine learning techniques and rule based techniques will be discussed in this article.

There is still potential for improvement in optimizing container scheduling based on fluctuating workloads while utilizing cluster resources. Based on this study and analysis the article has suggested a solution for Container Scheduling Optimization in Clustering Environment using a machine learning model with a rule-based approach as a hybrid approach to determine the scaling factor for the demand of the microservices. The suggested hybrid technique offers greater granularity in manipulating the scaling decisions. This is because the system predicts low-level parameters (demanding CPU Units) based on high-level data, rather than directly influencing the scaling factor.

This research aims to assess the reliability of the proposed autoscaling strategy by analyzing the results obtained. This offers an implementation of autoscaling using microservices and a reference design for domains that are shared by the community.

Keywords: microservices, container, machine learning, Fuzzy rules, control theory

TABLE OF CONTENT

LIST OF FIGURES

Figure **Description** **Page**

LIST OF TABLES

Table **Description** **Page**

LIST OF ABBREVIATIONS

Abbreviation	Description
ML	Machine Learning
SLA	Service Level Agreement
SLO	Service Level Objectives
VM	Virtual Machine
RAM	Random Access Memory
CPU	Central Processing Unit
QoS	Quality of Service
RIL	Reinforcement Learning
SVM	Support Vector Machine
SVR	Support Vector Regression
MAPE	Monitor Analyze Plan Execute
IDE	integrated development environment
SOA	Service Oriented Architecture
IoT	Internet of Things
PID	Proportional Integral Derivative
PI	Proportional Integral
I	Integral
ANN	Artificial Neural Network
RT	Regression Tree
RF	Random Forest
AWS	Amazon Web Services

LIST OF APPENDICES