A COST-EFFECTIVE AUTO-SCALING STRATEGY BETWEEN SERVER AND SERVERLESS ARCHITECTURES FOR CLOUD DEPLOYMENTS

LDSB Weerasinghe

(189357U)

Degree of Master of Science in Computer Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

May 2020

A COST-EFFECTIVE AUTO-SCALING STRATEGY BETWEEN SERVER AND SERVERLESS ARCHITECTURES FOR CLOUD DEPLOYMENTS

LDSB Weerasinghe

(189357U)

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

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

May 2020

DECLARATION

I declare that this is my own work and this 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 acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my 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:…….

Name: Sidath Weerasinghe

The supervisor should certify the dissertation with the following declaration.

The above candidate has carried out research for the Masters Dissertation under my supervision.

Signature of the supervisor: ……. Date: ……..

Name: Dr. Indika Perera

ABSTRACT

Nowadays, cloud computing is marking as an emerging technology in the universe. Most of people use cloud environments to deploy their systems. Many enterprise systems give their benchmarks according to the specific instances in the cloud. Cloud providers provide primary services as SaaS, PaaS, IaaS and FaaS, then Cloud consumers use those services as per their demands.

The systems which are developed on several years ago are still running upon on-premise environments. They are tightly coupled legacy systems. Hence people cannot adopt new technologies to those monolithic systems. Nevertheless, people who are in the software industry need to integrate new technologies into their systems to stay competitive in the IT industry. To get into the new technology that should be cost-effective, more reliable, and should need to cater the new and existing functionalities of the system. One of the problems is that migrating the whole system or partial system to the new technology is hazardous. Sometimes the system needs to handle substantial traffic loads, according to the business needs. In order to satisfy sudden traffic spikes, systems need to scale horizontally or vertically. If their application server is deployed upon the on-premise server, scaling is complicated, and if the application is deployed on the cloud, VMs scaling is possible, but it can be very costly.

In this research, the goal is to overcome these scaling problems. The research component presents the strategy to load balancing the traffic between the server and serverless function. The cost of the serverless function calculated only for the executes requests, so it is cheaper than running a new server. Therefore this hybrid strategy is a highly cost-effective way to scaling the servers and also proven that the proposed system is capable of load balancing the traffic according to the server loads with the analytical results. Other researches are proven that this research component is the right solution for ongoing industry problems.

Keywords: Cloud Computing, Serverless, Load Balancing

ACKNOWLEDGMENTS

My sincere appreciation goes to my family for the continuous support and motivation given to manage my MSc research work. I also express my heartfelt gratitude to Dr. Indika Perera, my supervisor, for the supervision and advice given throughout to this research period. I also thank my batch mates, for their untiring help to find research materials. I am also thankful to WSO2 Telco (Pvt) Ltd, for providing support to complete this research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

SaaS	Software as a Service
PaaS	Platform as a Service
FaaS	Function as a Service
IaaS	Infrastructure as a Service
NRPE	Nagios Remote Plugin Executor
AWS	Amazon Web Services
GCP	Google Cloud Platform
TPS	Transitions Per Second