

CLOUD BASED RESOURCE PREDICTION TOOL FOR DEPLOYMENT MIGRATIONS FOR DEVOPS

A.G.N.M. Ranaweera

(179347F)

M.Sc. in Computer Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

January 2020

CLOUD BASED RESOURCE PREDICTION TOOL FOR DEPLOYMENT MIGRATIONS FOR DEVOPS

A.G.N.M. Ranaweera

(179347F)

This dissertation submitted in partial fulfilment of the requirements for
the Degree of MSc in Computer Science specializing in Information
System Management

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

January 2020

DECLARATION

I declare that this is my own work and this MSc Thesis Project Report 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, 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.

A.G.N.M.Ranaweera

Date

I certify that the declaration above by the candidate is true to the best of my knowledge and that this project report is acceptable for evaluation for the MSc Thesis (CS6997).

Dr. Dulani Meedeniya

Date

Dr. Indika Perera

Date

ABSTRACT

Cloud computing is a promising paradigm for delivering computer services for industries, due to the advances in technologies such as virtualization, networking, web services, etc. As the cloud host providers offer an infinite amount of resources to their customers, the overhead of procuring traditional infrastructure resources has been reduced. In addition to the infinite resource offers, the dynamic resource scaling feature in the cloud allows the enterprise application owners to choose an appropriate pricing model. This feature in the cloud reduces the risk of over-provisioning and under-provisioning. Selecting an appropriate dynamic resource scaling model in the cloud platform is a tedious task as it depends on the accuracy of the forecasted resource provisioning values.

Based on this deliberate background, this research intends to model a POC (Proof of Concepts) for the future resource prediction in the cloud platform which assists the dynamic resource provisioning scaling. These resource provisioning predictions were done using the performance metrics gathered in the on-premises servers. The performance metrics gathered during 30 days were evaluated using the ARIMA (Autoregressive Integrated Moving Average) time series model. To improve the accuracy of the seasonality effects, the 30 days matrices have been gathered. Out of many ARIMA models, the best fitted ARIMA model was chosen based on the least RMSE (Root Mean Square Error).

The best fitted ARIMA model has predicted the future resource requirement of resources, by evaluating the gathered metrics. The model was trained based on the 75% data sets consisting of 22 days metrics and was tested using 25% data sets consisting of 8 days metrics. As a best practice, the data set splitting ratio has been selected as a 75:25 which consists of a higher number of data matrices for training. The resource provisioning model has been evaluated in the IaaS (Infrastructure as a Service) layer of the Amazon cloud. As the experimental cloud infrastructure, an EC2 (Elastic Cloud Compute) in the Amazon Web Services has been used. The observed values were evaluated using RMSE, MAPE (Mean Percentage Error), precision, recall, and F1 score factors as accuracy assessment criteria. The obtained accuracy percentages of CPU, RAM, Disk, and Network were respectively 84.07%, 93.13%, 90.47% and 71.16% which demonstrates a considerable accuracy for future forecast prediction. Further, visualization of the future forecasted resource provisioning results is provided in a dashboard. The developed dashboard has been evaluated through a survey comprised of a questionnaire that was distributed among a DevOps team which consists of ten members. The sample of the survey consisted of both technical and non-technical users who have 4 to 14 years of work experience in the IT industry. The overall average satisfaction rate of the survey was 4.1 out of 5, which was a considerably acceptable result. Based on the survey results and accuracy percentages of assessment criteria, it can be concluded that the research work has been successful and the proposed dashboard will be used for resource provisioning management.

ACKNOWLEDGEMENTS

I would like to express my great appreciation to my research supervisor Dr. Dulani Meedeniya and the M.Sc. research coordinator Dr. Indika Perera, for their guidance, enthusiastic encouragement and useful critiques of this research work.

My grateful thanks are also extended to Mr. Dinesh Asanka and Miss Inoka Loluwegoda for guiding me in the entire academic life journey. Special thanks go to the colleagues in my office, for their help and encouragement throughout my study. I also would like to precise my gratitude to the staff members of the CSE department for supporting me in the entire academic years.

I am grateful to all of those with whom I have studied during these two years. Since my heartiest thanks go to my M.Sc. colleagues who encouraged me during the studies. Specially Mr.Tiran Wijesekara, Mrs. Trisha Amarasekara, Mr.Ryan Benjamin, Mr. Vilochane Vidyarathne, Mr.Lakshan Costa and Miss. Nithila. Without them, I would not able successfully complete this. I would also like to express my gratitude to Mrs. Damayanthi Herath, Mrs. Shyama Ranasinghe, and Miss Niroshika Wickramarathane. I always knew that you believed in me and wanted to see the best for me.

Finally, my deep and sincere gratitude to my family for their continuous and unparalleled love, help and support. I would like to extend gratitude to my ever-loving mother, father, sister, and brother; whose love and direction are with me in whatever I achieved.

TABLE OF CONTENTS

DECLARATION	i
ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES	vi
LIST OF TABLES	vii
LIST OF EQUATIONS	viii
LIST OF ABBREVIATIONS	ix
1. INTRODUCTION	1
1.1. Background	1
1.1.1. Overview for Business Intelligence System	3
1.1.2. Overview of DevOps Practices	4
1.1.3. Overview of Cloud Migrations	5
1.2. Problem Statement	6
1.3. Motivation	7
1.4. Objectives	7
1.5. Outline	8
2. LITERATURE REVIEW	10
2.1. Introduction to DevOps	10
2.1.1. Benefits of DevOps	11
2.1.2. DevOps Process and Tools	11
2.1.3. DevOps and Cloud	11
2.2. Introduction to Cloud Migration	12
2.2.1. Business Model of Cloud	12
2.2.2. Cloud Migration Strategies – “6 R” Framework	14
2.2.3. Decision Support Tools for Cloud Migration	15
2.2.4. Prediction Models for Adaptive Resource Provisioning in the Cloud	16
2.3. Overview and Comparison of Cloud Migrations Tools	17
2.3.1. AWS Application Discovery Tools for Cloud Migration	18
2.3.2. Azure Data Migration Tools	18
2.3.3. IBM Cloud Migration Tools	19
2.4. Conclusions Based on Literature Review	20
3. METHODOLOGY	21
3.1. Chapter Overview	21
3.2. Overview of the Proposed Approach	21
3.3. Gathering the Historical Performance Metric Values of Individual Workload Time Series	24
3.4. Manipulating the Historical Performance Metric Values	25
3.4.1. Data Cleansing	25
3.4.2. Data Mapping	26
3.5. Analyzing the Gathered Workload Data Traces	27
3.6. Applied the Time Series ARIMA Model to Predict the Resource Provisioning.	30
3.7. Presenting the Resource Provisioning Outcome in the Dashboard	35
3.8. Implement the Identified Model Using Amazon EC2 as the Experimental Infrastructure	36
3.9. Summary	36

4.	IMPLEMENTATION	37
4.1.	Chapter Overview	37
4.2.	Gathering Historical Performance Metric Values	37
4.3.	Resource Provision Prediction	38
4.4.	Visualized the Forecasted Resource Provisioning in a Dashboard	41
4.5.	Implement the Forecasted Resource Provisioning in the Amazon EC2 Platform..	43
4.6.	Summary	45
5.	EVALUATION	46
5.1.	Chapter Overview	46
5.2.	Evaluate the Resource Provisioning.....	46
5.2.1.	MAPE.....	47
5.2.2.	Accuracy Percentage.....	47
5.2.3.	RMSE	47
5.2.4.	Precision.....	48
5.2.5.	Confusion Matrix.....	48
5.2.6.	Recall.....	48
5.2.7.	F1 Score.....	49
5.2.8.	Evaluations of the CPU	49
5.2.9.	Evaluations of the RAM.....	50
5.2.10.	Evaluations of the Disk.....	52
5.2.11.	Evaluations of the Network	53
5.3.	Evaluate the Dashboard.....	54
5.3.1.	Ease of Use	55
5.3.2.	Usefulness.....	55
5.3.3.	Overall Satisfaction	56
6.	CONCLUSION.....	57
6.1.	Chapter Overview	57
6.2.	Research Contribution	57
6.3.	Research Limitation and Future Work	58
	REFERENCES.....	60

LIST OF FIGURES

Figure 1.1: Overview of BI System, Source: [8]	3
Figure 1.2: Overview of the DevOps Practices, Source: [11].....	4
Figure 1.3: Market Share of the Cloud Vendors (the year 2019), Source: [16].....	6
Figure 2.1. DevOps Processes, Source: [20].....	10
Figure 2.2. Cloud Computing Architecture, Source: [26].....	13
Figure 3.2: High-level Server Architecture of the BI System.....	22
Figure 3.1: Flow Chart of the Methodology for Resource Provisioning	22
Figure 3.3: The Process of Data Mapping	26
Figure 3.4: The Line Chart for the Resource Performance Metrics.....	28
Figure 3.5: Five Numbers Summary of a Box Plot.....	28
Figure 3.6: The Box Plot Chart for the Resource Utilization Data Traces	29
Figure 3.7: The seasonality differences of the current data set - CPU.....	30
Figure 3.8: Autocorrelation and Partial Autocorrelation Plots	32
Figure 3.9: Predicted values vs. Expected values	35
Figure 4.1: Data set Validation Code Snipped.....	39
Figure 4.2: Prediction Class Code Snipped.....	40
Figure 4.3: Data Uploading Interface in Quick Sight	41
Figure 4.4: Resource Provisioning Dashboard Summary	42
Figure 4.5: Amazon AMI Interfaces	43
Figure 4.6: Virtual Machine Parameters	44
Figure 4.7: PRP-Server Details	44
Figure 4.8: DB Instance Details	44
Figure 5.4: Overall Evaluations from Users	56

LIST OF TABLES

Table 2.1: Features and Limitation of Cloud Migrations Tools	19
Table 2.2: Comparison of Cloud Migrations Tools, Source [35]	20
Table 3.1: Season Identified in CPU Data Traces.....	30
Table 3.2 : RMSE Validations Results	33
Table 3.3 : Predicted values vs. Expected values.....	34
Table 4.1: Usage of Python libraries, Source [47]	39
Table 5.1: Confusion Matrix - Outline.....	48
Table 5.2: Confusion Matrix - CPU.....	50
Table 5.3: Eavluation Summary - CPU.....	50
Table 5.4: Confusion Matrix - RAM.....	51
Table 5.5: Eavluation Summary – RAM.....	51
Table 5.6: Confusion Matrix - Disk	52
Table 5.7: Eavluation Summary - Disk.....	53
Table 5.8: Confusion Matrix - Network.....	53
Table 5.9: Eavluation Summary - Network	54
Table 5.2: Questionnaires Form Template for Ease of Use	55
Table 5.3 : Questionnaires Form Template for Usefulness	55
Table 5.4: Results of the Questionnaires.....	56
Table 6.1: Evaluation Summary for Resources.....	58

LIST OF EQUATIONS

Equation 1: MAPE	47
Equation 2: Accuracy Percentage	47
Equation 3: RMSE	47
Equation 4: Precision	48
Equation 5: Recall	49
Equation 6: F1 Score	49

LIST OF ABBREVIATIONS

Abbreviation	Description
API	Application Program Interface
ARIMA	Autoregressive Integrated Moving Average
AWS	Amazon Web Services
BA	Business Analysis
BI	Business Intelligence
DB	Data Base
Dev	Development
DevOps	Development and Operations
EA	Enterprise Application System
ETL	Extraction, Transformation, Loading
FTE	Full Time Equivalent
IaaS	Infrastructure as a Service
IT	Information Technology
MAPE	Mean Percentage Error
ML	Machine Learning
PaaS	Platform as a Service
POC	Proof of Concept
QA	Quality Assurance
RMSE	Root Mean Square Error
SaaS	Software as a Service
SD	Standard Deviation
SSIS	SQL Server Integration Server
VM	Virtual Machine