LEARNING-CENTRIC RESOURCE MANAGEMENT IN ENERGY-EFFICIENT CLOUD COMPUTING

B.A.J Pushpakumara

219386U

Degree of Master of Science in Computer Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

June 2024

LEARNING-CENTRIC RESOURCE MANAGEMENT IN ENERGY-EFFICIENT CLOUD COMPUTING

B.A.J Pushpakumara

219386U

Thesis/Dissertation submitted in partial fulfilment of the requirements for the degree of MSc in Computer Science specializing in the Software Architecture

Department of Computer Science

University of Moratuwa
Sri Lanka

June 2024

DECLARATION

I declare that this is my own work and this thesis/dissertation 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 the whole or part of future works (such as articles or books).

Signature: Date: 2024/06/29

The above candidate has carried out research for the PhD/MPhil/Masters thesis/dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Prof I Perera

Signature of the Supervisor: Date: 29/06/2024

ACKNOWLEDGEMENT

I extend my sincere gratitude to all the individuals who have directly or indirectly helped me in completing this project report. I would like to express my heartfelt thanks to Professor Indika Perera, my supervisor, for his invaluable guidance, encouragement, and support throughout the entire research process. I would also like to acknowledge my wife and family for their unwavering support and encouragement. I am also grateful to my colleagues and friends who have provided me with valuable feedback and inspiration. I would like to express my thanks to the resources and facilities provided by my university, which have facilitated my work. Finally, I would also like to extend my sincere appreciation to my workplace for their invaluable support in managing my research work. Their constant encouragement and resources have made it possible for me to pursue my goals.

ABSTRACT

Cloud computing has revolutionized the IT landscape by introducing a subscription-based model for computing resources. This on-demand resource delivery, coupled with a pay-as-you-go pricing structure, empowers users with unparalleled convenience and elasticity in resource scaling. This convenience has led to a rapid growth in cloud data centers, which are predicted to consume a significant portion of the world's electrical usage by 2025 [1]. This growth poses a substantial energy challenge, highlighting the need for energy-efficient cloud computing solutions [3].

Modern cloud data centers are complex, comprising heterogeneous servers, networking elements, power distribution systems, and cooling systems, all interacting in complex ways [16]. This complexity, coupled with the exponential growth of data centers and diverse workloads, necessitates research in energy-efficient cloud computing.

Dynamic VM consolidation plays a significant role, Among various energy-saving techniques employed in cloud data centers, which includes migration of virtual machines (VMs) between servers to consolidate workloads and idle servers, thereby saving energy. However, existing dynamic selection algorithms for VM migration are heuristic-based and lack accurate utilization predictions, hindering their effectiveness.

This research presents a novel approach to dynamic VM consolidation, leveraging a Machine Learning (ML) driven model. This model forecasts CPU utilization trends for individual VMs by analyzing their historical data. A VM selection algorithm is then proposed, utilizing the ML model to perform intelligent dynamic VM consolidation. The proposed algorithm's performance is assessed using a simulated cloud data center driven by real cloud workload traces, focusing on energy efficiency and customer Service Level Agreements (SLAs) violations, compared against existing heuristic-based algorithms.

TABLE OF CONTENT

DECLARATION 2

ACKNOWLEDGEMENT 3

ABSTRACT 4

TABLE OF CONTENT 5

LIST OF FIGURES 8

LIST OF TABLES 9

LIST OF ABBREVIATIONS 10

Introduction 11

1.1 Introduction 12

1.1.1 Limitations of Existing Techniques and the Rise of VM Consolidation 14

1.1.1.1 Challenges with Other Server Components. 14

1.1.1.2 Dynamic VM Consolidation as a Solution 14

1.1.2 Balancing Energy Efficiency with Performance 15

1.2 Motivation 16

1.3 Research Problem 16

1.4 Research Objectives 17

1.5 Structure of the thesis 17

Literature Review 19

2.1 Breaking Down Energy Consumption in Cloud Environments 20

2.2 The Environmental and Economic Implications of High Energy Consumption 22

2.2.1 Economic Burden and Resource Underutilization 23

2.2.2 Environmental Consequences and CO2 Emissions 23

2.2.3 Challenges of Heat Dissipation and Server Density 23

2.2.4 Focus on Efficiency and Sustainability 24

2.3 Existing Techniques for Enhancing Energy Efficiency in Cloud Computing 24

2.3.1 Hardware Techniques 25
2.3.1.1 Dynamic Voltage and Frequency Scaling 25
2.3.1.2 Energy-Efficient Data Center Design 27
2.3.2 Virtualization Techniques 28
2.3.2.1 Load Balancing 29
2.3.2.2 Resource Over-commitment 30
2.3.2.3 Practices of Workload Scheduling and Provisioning 31
2.3.2.4 Dynamic Consolidation of Virtual Machines 32
2.4 Exploration of Virtual Machine Consolidation Studies 34
2.5.1 VM Allocation Policies 35
2.5.2 VM Selection Policies 35
2.5.2.1 Minimum Utilization (MU) 36
2.5.2.2 Maximum Correlation (MC) 37
2.5.2.3 Random Selection (RS) 38
2.5.2.4 Minimum Migration Time (MMT) 39
2.5 Other Runtime Management and Optimization Techniques 40
Methodology 41
3.1 CPU Utilization's Impact on Energy Usage 42
3.2 VM selection policy 43
3.3 Development of the Machine Learning Model 46
3.3.1 High level steps involved. 46
3.4 Deployment of the Machine Learning Model 49
3.5 System Evaluation 49
Implementation 50
4.1 CloudSim Framework 51
4.1.1 The primary components of the CloudSim framework 51
4.2 VM selection policy implementation 52
4.3 Machine Learning Model implementation and training 55
4.4 Experimental Setup and testing of the proposed model. 56

Evaluation 59

5.1 Statistical Effectiveness of the model 60

5.2 Heuristic MU vs. ML VM Selection against single data set...60

5.3 Heuristic MU vs. ML VM Selection against multiple data sets. 61

Conclusion 64

6.1 Research Contribution 65

6.2 Study limitations 66

6.2.1 Limitations Due to Assumptions and Factors Ignored 66

6.2.2 Limitations of the CloudSim Simulation Framework..67

6.3 Future Directions 68

References 70

LIST OF FIGURES

Figure 1 - Best and expected global electricity consumption from [1] 12
Figure 2 - The global power repartition trends for ICT [1] 13
Figure 3 - Proportionality of Power Consumption in a Data Center [30,31] 22
Figure 4 - Shutting down the underutilized host servers [4] 33
Figure 5 - Minimum Utilization (MU) VM Selection policy algorithm. 36
Figure 6 - Maximum Correlation VM Selection Algorithm 37
Figure 7 - Random Selection VM selection policy Algorithm. 38
Figure 8 - Minimum Migration Time VM selection Policy Algorithm 39
Figure 9 - Power model defined [3] 42
Figure 10 - Total energy model defined [3] 43
Figure 11 - The Architecture of a cloud data center [16] 44
Figure 12 - ML-driven VM consolidation System Diagram 45
Figure 13 - Actual CPU utilization of a cloudlet over time. 48
Figure 14 - Actual and Predicted CPU utilizations over the time. 48
Figure 15 - Abstract PowerVmSelectionPolicy implementation. 53
Figure 16 - ML VM Selection Algorithm. 54
Figure 17 - PowerVmSelectionPolicyMachineLearning implementation. 54
Figure 18 – Energy consumption with LR MU (kWh) and LR ML (kWh) 63
Figure 19 – VM migration count with LR MU (kWh) and LR ML (kWh) 63

LIST OF TABLES

Table 1 – ML Model statistical comparison 46
Table 2 – lr_mu vs lr_ml simulation output comparison 61
Table 3 – Summarized simulation results for different data sets 62

LIST OF ABBREVIATIONS

VM	Virtual Machine
ML	Machine Learning
CPU	Central Processing Unit
DRAM	Dynamic Random Access Memory
RAM	Random Access Memory
QoS	Quality of service
ICT	Information and Communications Technologies
DVFS	Dynamic Voltage and Frequency Scaling
SLA	Service Level Agreement
AC	Alternating Current
DC	Direct Current
CMOS	Complementary Metal-Oxide-Semiconductor
UPS	Uninterruptible Power Supplies
DCIM	Data Center Infrastructure Management
OS	Operating System
LR	Local Regression
MU	Minimum Utilization
MMT	Minimum Migration Time
RS	Random Selection
ARIMA	Auto Regressive Integrated Moving Average
MC	Maximum Correlation
DRL	deep reinforcement learning
DPM	Distributed Power Management
TCO	Total Cost of Ownership
TCA	Total Cost of Acquisition