Faculty of Information and Communication Technology

ENHANCED MODEL TO MINIMIZE FUTURE DOWNTIME: CASE STUDY OF MALAYSIA CLOUD PROVIDERS TOWARDS NEAR-ZERO DOWNTIME

Clarissa Terry Clement

Master of Computer Science

(Database Technology)

2015

ENHANCED MODEL TO MINIMIZE FUTURE DOWNTIME: CASE STUDY OF MALAYSIA CLOUD PROVIDERS TOWARDS NEAR-ZERO DOWNTIME

CLARISSA TERRY CLEMENT

A project submitted
in fulfillment of the requirements for the degree of Master of Computer Science
(Database Technology)

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2015

DECLARATION

I declare that this project entitled "Enhanced Model to Minimize Future Downtime: Case Study of Malaysia Cloud Providers Towards Near-Zero Downtime" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : …….

Name : Clarissa Terry Clement

Date : 6th July 2015

APPROVAL

I hereby declare that I have read this report and in my opinion this report is sufficient in terms of scope and quality as a partial fulfillment of Master of Computer Science (Database Technology).

Signature : ……

Supervisor Name : P. M. NORHAZIAH BINTI MD SALLEH

Date : 6th July 2015

DEDICATION

This thesis is dedicated to my children, Neo Elleandro and Mia Avila. I give my deepest expression of love and appreciation for being there for me throughout the entire master program. Both of you have been my best cheerleaders.

I also dedicate this thesis and give special thanks to my best friends, Khairil Ashriq and Paul Major who have supported me all the way since the beginning of my studies.

ABSTRACT

In providing tremendous access to data and computing power of thousands of commodity servers, large-scale cloud systems must address a new challenge: they must detect and recover from a growing number of failures, in both hardware and software components. The growing complexity of technology scaling, manufacturing, design logic, usage, and operating environment increases the occurrence of failures. Unfortunately, downtime handling has proven to be problematic in today's cloud systems. The downtime recovery path is often complex, under-specified, and tested less frequently than the normal path. As indicated by recent cloud outage incidents, existing large-scale cloud systems are still fragile and error-prone. The purpose of this study to identify the issues causing cloud downtime, to investigate the recovery ability of the database during cloud downtime and to propose enhanced model that can be used to minimize the future downtime.

ABSTRAK

Dalam penyediaan akses kepada data yang besar dan kuasa ribu pelayan komoditi pengkomputeran, sistem pengkomputeran awan berskala besar perlu menangani cabaran baru: mereka harus mengesan dan membaik pulih keadaan yang semakin bermasalah, dalam kedua-dua perkakasan dan perisian komponen. Kerumitan yang semakin meningkat dalam skala teknologi, pembuatan, reka bentuk logik, penggunaan dan sistem operasi telah meningkatkan lagi kegagalan. Malangnya, pengendalian gangguan telah terbukti mengalami masalah dalam pengkomputeran awan hari ini. Jalan pemulihan sistem gangguan adalah kompleks, dan kurang diuji. Seperti yang ditunjukkan oleh baru-baru ini insiden gangguan pengkomputeran awan, pengkomputeran awan masa kini ada masih belum stabil dan kesilapan yang sering berlaku. Tujuan kajian ini untuk mengenal pasti isu-isu yang menyebabkan gangguan pengkomputeran awan, untuk menyiasat keupayaan membaik pulih pangkalan data semasa gangguan pengkomputeran awan dan untuk mencadangkan model yang telah dipertingkatkan yang boleh digunakan untuk mengurangkan gangguan dimasa depan.

ACKNOWLEDGEMENTS

First and foremost, I would like to take this opportunity to express my sincere acknowledgement and gratitude to my supervisor Professor Madya Norhaziah Bt. Md. Salleh from the Faculty of Information and Communication Technology Universiti Teknikal Malaysia Melaka (UTeM) for her immeasurable amount of support, guidance and encouragement towards the completion of this thesis.

I would also like to express my appreciation to Dr. Mohd Sanusi Bin Azmi and Dr. Asmala Bin Ahmad from Faculty of Information and Communication Technology, panels of this project for their advice and suggestions in giving the suitable tittle and research methodology on completing this research.

Particularly, I would also like to give credit to the cloud providers in Malaysia for cooperating in my industry survey. Their feedback plays important part in this thesis.

Special thanks to all my peers, my family and friends for their moral support in completing this degree. Lastly, thank you to everyone who had been to the crucial parts of realization of this project.

TABLE OF CONTENTS

		PAGE
DECLARATION		
DEDICATION		
ABSTRACT		**i**
ABSTRAK		**ii**
ACKNOWLEDGEMENTS		**iii**
TABLE OF CONTENTS		**iv**
LIST OF TABLES		**vi**
LIST OF FIGURES		**vii**
LIST OF APPENDICES		**viii**
CHAPTER		
1.	**INTRODUCTION**	**1**
1.0	Background	1
1.1	Background of Project	4
1.2	Problem Statement	5
1.3	Research Questions	6
1.4	Objectives	7
1.5	Significant and Research Contribution	7
1.6	Research Scope and Limitation	7
1.7	Organization of the Report	8
2.	**LITERATURE REVIEW**	**10**
2.1	Introduction	10
2.2	Related Works	11
2.3	Comparison Between Related Works	22
2.4	Theoretical Framework	25
2.5	Summary	26
3.	**RESEARCH METHODOLOGY**	**27**
3.1	Introduction	27
3.2	Research Design	28
3.2.1	Research process design	29
3.3	Research Plan	30
3.3.1	Phase 1	31
3.3.2	Phase 2	31
3.3.3	Phase 3	32
3.4	Data Collection	32
3.5	Question Design	33
3.6	Data Analysis	34
3.7	Summary	35

4. IMPLEMENTATION AND DATA ANALYSIS 36
4.1 Introduction 36
4.2 Questionnaire design 36
4.2.1 Determine the questions to be asked 37
4.2.2 Question type 38
4.2.3 Design and construct the questionnaire 38
4.3 Data Analysis 39
4.4 Summary 43

5. FINDINGS AND DISCUSSION 44
5.1 Introduction 44
5.2 Findings 44
5.3 Discussion and proposed model 48
5.4 Summary 52

6. CONCLUSION 53
6.1 Introduction 53
6.2 Conclusion 53
6.3 Limitation of research 54
6.4 Future work 54

REFERENCES 55
APPENDICES 61

LIST OF TABLES

TABLE	TITLE	PAGE
2.1	Comparison between related of works (literature review)	22
3.1	Summary of the research method	29
4.1	The process to generate questions	37

LIST OF FIGURES

FIGURE	TITLE	PAGE
1.1	Structure of cloud computing model.	2
2.1	Pre-copy approach	12
2.2	Post-copy Approach	14
2.3	Resource Reservation Based Live Migration of Multiple Virtual Machines	17
2.4	PipeCloud's implementation components	19
2.5	Cross-layer control and information flow for autonomic management in a virtualized datacenter	21
2.6	Theoretical framework of research	25
3.1	Research Process in the study	30
3.2	Research Plan	31
4.1	Types of downtime	40
4.2	Number of downtime event yearly	40
4.3	Duration downtime	41
4.4	Root cause cloud downtime	42
4.5	Strategies and technologies to ensure system availability and data protection	42
5.1	PipeCloud's implementation components	51

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	Questionnaire for Cloud Downtime Recovery Survey	62
B	Response Summary	65

CHAPTER 1

INTRODUCTION

1.0 Background

The emergence of cloud computing an essential change in the way information technology (IT) administrations are produced, kept up, utilized and paid for. Data innovation administrations inside of an association have turned out to be considerably more perplexing as of late, creating the administration and the circulation of processing assets to be at such a complicated level, that the product included had made registering more extravagant than any other time in recent memory. On the other hand, the guarantee of distributed computing has permitted associations to show these registering assets to customers in a manner that generously helps associations lessen the forthright expenses of figuring; in this way, better adjust their needs and spending plans (Mikkilineni and Sarathy, 2009). In distributed computing, moving information into the cloud offers incredible settlement to clients since they don't need to inconvenience about the intricacies of direct equipment administration (V. Krishna Reddy et al., 2011). This all-new business readiness empowers the conveying and handling of new items to reach skillful levels notwithstanding amid top times consistently. The cloud turns into a versatile and charming system when effectively actualized by associations, definitely pulling at last clients for their individual needsThe cloud can be conveyed in three models which have diverse elements and ways as shown in figure 1.1 (Dr Nashaat el-Khameesy et al., 2012).

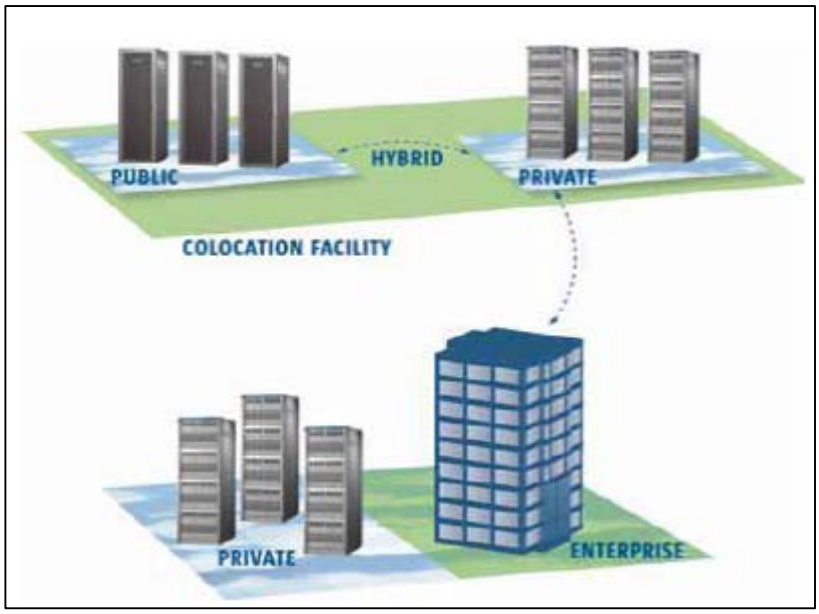

Figure 1.1: Structure of cloud computing model (Dr Nashaat el-Khameesy et al., 2012).

i. Public Cloud

A public cloud is one in which the services and foundation are given off-site over the internet. These clouds offer the best level of proficiency in shared assets; nonetheless, they are more exposed than private clouds. Open clouds are managed by third party, and applications from diverse clients are prone to be combined on the cloud's servers, storage systems, and networks (Dr Nashaat el-Khameesy et al., 2012).

ii. Private Cloud

A private cloud is one in which the services and framework are well-kept on a private network. These clouds offer the best level of security and control, however they oblige the organization to still buy and keep up all the software and infrastructure, which diminishes the expense investment funds (Dr Nashaat el-Khameesy et al., 2012).

iii. Hybrid Cloud

A hybrid cloud combines multiple clouds (private, community of public) wherever those clouds retain their unmatchable identities, however square measure limit along as a unit. A hybrid cloud might pass standardized or possession access to knowledge and applications, still as application movableness.

Cloud provides services that can be categorized into three groups:

i. Software as a Service (SaaS):

This model gives access to programming applications over a system. Accessible to different clients over the Internet, this model backings applications with their own server farm which permits them to basically keep running on the cloud, killing the need to introduce and run the application on a PC. There are numerous buyer based SaaS administrations, for example, facilitating and stockpiling administrations; Dropbox being the most evident stockpiling administration.

ii. Platform as a Service (PaaS)

PaaS is for software suppliers who need to focus on the improvement of new applications without the expense and multifaceted nature of purchasing and dealing with the equipment and programming required. PaaS is cloud administrations with inherent incorporation of web administrations and databases; so essentially designers utilize these administrations to build up their own item from a current SaaS or add to another web application.

iii. Infrastructure as a Service (IaaS)

This represents hardware services such as stockpiling, force and memory administrations are likewise given. The best thing about this model is clients can pay-as-you-go, in a manner of speaking. There is likewise the choice of marking an agreement for a particular measure of time. This is primarily in light of the fact that the client is in charge of everything, significance the clients need to set up the application environment in their virtual machines.

Meanwhile, there are a considerable measure of ventures to be taken for the buildup to encompass this advancement to develop into an enduring achievement. Numerous regulations must be placed set up, the interoperability of effectively existing cloud situations should be created, and above all the adaptability of the innovation needs to keep on growing to help the customary IT situations which have upheld us for a considerable length of time. The key attributes and variables that these associations would use to figure the conceivable achievement of distributed computing selection is still, some-what, not yet decided. A few inquiries are as yet being raised about the accessibility of cloud stages, as no association needs to win a terrible notoriety as a result of downtime.

1.1 Background of Project

Cloud computing has become a new trend of advancement in the world of information technology today in which information technology resources are delivered as a service via the internet. From this, it is assumed that it is the newest internet-based technology. The increasing growth of this new technology coupled with it imminent productivity and benefit has made most organizations turn to the cloud. The reason being that most organizations can now deploy and

manage their IT services via a virtual machine in the cloud, this reduces the enormous cost being spent on setting up, managing and maintaining previous local systems and infrastructure.

Among the numerous and different concerns, the risk of inadequate service availability has been identified as one of the top obstacles to adoption of Cloud computing (Li et al. 2013). IT managers hate system downtime, but the harsh reality is that even the best plans and preparation cannot prepare for every circumstance, and even the simplest oversights can snowball into serious events that are difficult and costly to remediate. Given the reality that it is hard to eliminate the downtime of the data center, systems behind cloud services most of the current studies emphasized data/system backup from the perspective of Cloud customers (S. Bigelow, 2011). For example, employing multiple cloud providers was suggested as being a plausible solution to very high availability service delivery (Armbrust et al. 2010). With this point of view, this research would explore ways to minimize cloud downtime in the future.

1.2 Problem Statement

Many researchers concentrated on the security and privacy issues of cloud computing and came up with several methods to enhance the data storage security in the cloud. However, not many researchers take any account on handling the issues of cloud downtime. Despite the fact that distributed computing administration suppliers touted the security and unwavering quality of their administrations, genuine sending of distributed computing administrations is not as sheltered and solid as they claim (Chen et al., 2012). In 2009, the significant distributed computing sellers progressively seemed a few mischances. Amazon's Simple Storage Service was interfered with twice in February and July 2009 (State et al., 2012). This mishap brought about some system locales depending on a solitary kind of capacity administration were

compelled to a stop. In March 2009, security vulnerabilities in Google Docs occasion prompted genuine spillage of client private data. Google Gmail additionally showed up a worldwide disappointment up to 4 hours (Vali et al., 2012). It was uncovered that there was not kidding security weakness in VMware virtualization programming for Mac form in May 2009. Individuals with ulterior thought processes can exploit the defenselessness in the Windows virtual machine on the host Mac to execute pernicious code. Microsoft's Azure distributed computing stage likewise occurred a genuine blackout mischance for around 22 hours (S. Deepak 2013). Genuine security episodes even prompt breakdown of distributed computing merchants. As overseers' abuse prompting loss of 45% client information, distributed storage merchant LinkUp had been compelled to close.

The cases showed that the issue of cloud downtime is not fully understood. The root cause of the outage is yet to be identified while the recovery of the database is still an unsolved issue where it led to hours even days of system downtime.

Although there is no absolute means for preventing outage (Bigelow 2011), it is still worthwhile for cloud vendors to learn from the existing mistakes, so they are able to minimize the risk of future downtime.

1.3 Research Questions

In the exploration of the early stage of the study, some research questions arises which motivated the progress of the study itself. The following are the research question of the study:

i. What type of downtime does the company experienced?

ii. How often does the downtime occur?

iii. What is the cause of the downtime?

iv. How long is the duration of the downtime?

v. What are the actions taken by the company to ensure system availability and data protection?

1.4 Objectives

The following are objectives of the study:

i. To identify the issues causing cloud downtime.

ii. To investigate the recovery ability of the database during cloud downtime

iii. To propose enhanced model to minimize the future downtime

1.5 Significant and Research Contribution

In this study, it has been recognized that zero downtime is impossible for large-scale Internet services. By learning from the previous and others' mistakes, nevertheless, it is possible for cloud providers to at least keep the downtime short. To facilitate the research on investigating the database recovery ability, this study will perform an industry survey of how the Malaysia cloud providers manage their system downtime. In addition to a set of findings, the research will be able to generate a theoretical framework by classifying the downtime root causes and propose suitable model to manage the issue. By reviewing the literature and collecting data from the industry survey, the proposed enhanced model may be adapted in cloud services and can be used the providers in the future.

1.6 Research Scope and Limitation.

i. Users of cloud computing will be able to access their data while the downtime.

ii. Cloud providers will be able to use the proposed enhanced model to minimize the downtime event.

The research on how to minimize the downtime issue on cloud computing has not been widely explored. Most of the literature reviews are concerning the method of virtual machine migration on how the total time migration and downtime can be reduced. By exploring the methods, we may be able to adapt them in the cloud services.

1.7 Organization of the Report.

This report consists of six chapters and structured as follow:

Chapter 1: Introduction

Chapter 1 is the beginning part of this study and contains the introductory. This chapter provides information about the origins of the research. In introduction chapter, the main points explain about a brief overview of background study, problem definition, research questions, the objectives of the study, significant and research contribution, and lastly the scope and limitation of the study.

Chapter 2: Literature Review

Chapter 2 is the literature review. It will explain about the related works by previous researchers using several technique to reduce system downtime. Then, the literature will focus on the comparison techniques.

Chapter 3: Research Methodology

Chapter 3 discussed about on the methodologies that utilized as a part of this study to accomplish the research objectives. It includes research design, research plan, data collection, question design and data analysis.

Chapter 4: Implementation and Data Analysis

The fourth chapter discussed on how the questionnaire is designed and the analysis of the data collected from the industry survey.

Chapter 5: Findings and Discussion

Chapter 5 discussed about the study's findings, and some discussion regarding the result that reflects the objectives.

Chapter 6: Conclusion

Chapter 6 gives a summary of this research and study's limitations as well as routes for future research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The importance of cloud computing is increasing and it is receiving a growing attention in the scientific and industrial communities (Hashizume et al., 2013). The main concept of cloud computing is to improve the capability of cloud in order to reduce the computing load of client side. The ultimate goal is to simplify the client side to become the input device, which can take advantage of the powerful computing capability of Cloud (Dr Nashaat el-Khameesy et al., 2012). Taking into account the services provided by cloud computing is numerous but still there are some security concerns that are to be redressed, especially because cloud users have no choice but to rely on the service provider (Kaur and Bhardwaj, 2012).

Numerous critical system rely on upon cloud services. The U.S. military is allowing officers with access to military-run distributed computing for "discriminating observation and choice making data" (Ellen Messmer, 2010). Instructive foundations are utilizing Facebook and Twitter warnings to show crisis alarms. In 2011, cloud downtime has accentuated the clients dependence on distributed computing. On the morning of April 21, Amazon's Northern Virginia server farm began having network and disappointment issues with its Elastic Block Store (EBS), and Elastic Compute Cloud (EC2). While the vast majority of the issues were determined inside of 24 hours, the complete recuperation took until April 25. This downtime has influenced online locales, for example, Foursquare, Reddit and Quora, which rely on upon Amazon's cloud

(Amazon, 2011 and S. Lohr, 2011). The root cause of the outage was traced to an operator error during a routine upgrade meant to increase capacity (Amazon, 2011).

These discriminating administrations are driving the interest for changes in cloud accessibility. Accessibility is the division of time that a framework is prepared to give right administration and does not experience arranged or impromptu downtime. Accessibility can be liable to contractual commitments; specifically, Service Level Agreements (SLA) for business applications oblige that cloud administration and applications be exceptionally accessible. For instance, the Google Apps for Business SLA stipulates that applications be accessible 99.9% of the time, generally Google is at risk to repay clients for inaccessibility (Google, 2011).

2.2 Related Works

Many companies, public as well as private realize the operational inefficiencies of their data centers which sways both their working costs notwithstanding gigantic capital consumptions. With figuring accessible as an administration in a server farm virtualization worldview, numerous organizations think about virtualizations as an answer for obtain administrations. In this manner, more applications will be relocated to the virtual machine (VM). Virtual machine relocation is a standout amongst the most essential advantages of server virtualization innovation. The capacity to relocate a virtual machine starting with one physical host then onto the next can essentially help an association's fiasco recuperation endeavors and enhance business readiness. It likewise proves to be useful when an executive needs to close down a physical server for upkeep or updates in light of the fact that server downtime no more equivalents application downtime. There are few techniques proposed by previous researchers to reduce the downtime event during the migration process.