UNIVERSITI PUTRA MALAYSIA

ENHANCED REPLICATION STRATEGY WITH BALANCED QUORUM TECHNIQUE AND DATA CENTER SELECTION METHOD IN CLOUD ENVIRONMENT

FAZLINA BINTI MOHD ALI

FSKTM 2022 13

ENHANCED REPLICATION STRATEGY WITH BALANCED QUORUM TECHNIQUE AND DATA CENTER SELECTION METHOD IN CLOUD ENVIRONMENT

By

FAZLINA BINTI MOHD ALI

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

March 2022

COPYRIGHT

DEDICATION

This thesis is dedicated to my beloved father, mother, husband, sons, mother in-law, brother, sisters, family members, friends and everyone that involved in this journey.

Thank you for all the prayers.

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

ENHANCED REPLICATION STRATEGY WITH BALANCED QUORUM TECHNIQUE AND DATA CENTER SELECTION METHOD IN CLOUD ENVIRONMENT

By

FAZLINA BINTI MOHD ALI

March 2022

Chairman : Associate Professor Rohaya Latip, PhD
Faculty : Computer Science and Information Technology

Digital data are growing tremendously, and the massive amount of data are core contributors to data evolutions across the globe. In this Industry Revolution (IR) 4.0 era, heterogenous data are generated in various sources and platforms. In order to store these huge volumes with heterogenous data categories, cloud computing became the mainstream solution to provide multiple services to keep safe, process and distribute the data. As each data is substantial to everyone, cloud computing emphasises and facilitates a fast and flexible platform to users. Despite continually standing as a resilient storage provider to heterogenous data, the prominent issue encountered as performance challenges is to accommodate sufficient services with high data cloud storage. The issues are crucial because poor data availability and accessibility are often influencing delays on data retrieval which ultimately leads to system performance degradations.

In order to mitigate the issues, 'cloud data replication' is commonly implemented for better data performance and promising business continuity. Cloud replication is recognised as storing more than one copy of data in multiple distributed storage nodes. Additionally, cloud replication is well-known as a comprehensive technique which is capable in serving high data availability, faster response time, and better fault tolerance yet cost-effective for both cloud users and provider. Regardless of the replication valuable keys, occasionally this technique has the tendency to attain performance degradation issues too. Some existing research works did not considers substantial factors in determining popular files. As a result, the poor popular file selection strategies affect file download and lead to ineffective response time. Besides, there are also issues that arise during data placement techniques consuming extensive space and high replication time due to replica copies located in every storage node in the same cluster environment to achieve high data availability. Furthermore, there are also problems identified when choosing an appropriate data center to store replica copies caused by ineffective data center selection criteria. Holistically, there are many previous studies which have

enlightened issues in cloud replications such as delayed response time, high replication time, extensive storage space, and massive network consumption due to inefficient replication strategies implemented in cloud environments.

Therefore, to evade the negative probabilities and satisfy users, an established and extensive cloud replication strategy is significant to be deployed in a cloud environment to enhance the overall replication performance.

In this research, there are three (3) main contributions proposed for a cloud replication environment. Firstly, this research proposed an Enhanced Replication Strategy (ERS) that improves response time for file download in a replication cloud environment without neglecting the popular file selection. Secondly, this research proposed a dynamic Balanced Quorum (BQ) technique for replica placement to reduce storage consumption without disregarding a faster replication time and preserve data availability. As for the third contribution, an efficient Data Center Selection Method (DCSM) was proposed to ensure files are available in the best local data center; thus, this research is able to facilitate efficient network usage and minimise replication frequency in a cloud replication environment.

The proposed ERS, BQ and DCSM are agile depending on the replication environment requirements and dynamically adapt user access patterns, guarantee that essential data is always available and accessible by users. Thorough experiments were conducted using the 'CloudSim' simulation tool. The dataset generated by CloudSim was in random structured master files in various sizes. The simulation results were analysed and the analytical graphs are presented in the discussions. In order to validate the competence of these proposed algorithms, results were compared with another similar research work known as DPRS algorithm. Apparently, the proposed ERS, BQ and DCSM in this research outperformed the DPRS, concurrently evidenced 10.47% improvement in average response time, reduced storage consumptions are 11.43% and 31%, accelerated replication time with 6% and 20%, sustained high data availability between 98.59% to 99.03%, offered effective network usage reduced by 20% more efficiency and finally betterments on overall replication frequency with 14%. Individual measurement metrics in this research was outpaced the other existing method, the DPRS and contributed better performance in cloud replication environment.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

STRATEGI REPLIKASI YANG DITINGKATKAN DENGAN KAEDAH PEMILIHAN TEKNIK QUORUM DAN KAEDAH PEMILIHAN PUSAT DATA DALAM PERSEKITARAN PENGKOMPUTARAN AWAN

Oleh

FAZLINA BINTI MOHD ALI

Mac 2022

Pengerusi : Profesor Madya Rohaya Latip, PhD
Fakulti : Sains Komputer dan Teknolologi Maklumat

Data digital berkembang dengan pesat dan jumlah data yang besar adalah penyumbang utama kepada evolusi data di seluruh dunia. Dalam era Revolusi Industri (IR) 4.0 ini, data heterogen dijana dalam pelbagai sumber dan platform. Untuk menyimpan jumlah data yang besar dengan kategori data yang pelbagai, pengkomputeran awan menjadi penyelesaian dengan menyediakan pelbagai perkhidmatan untuk memastikan keselamatan memproses dan mengedarkan data. Oleh kerana setiap data penting kepada semua, pengkomputeran awan menekankan penyediaaan platform yang cepat dan fleksibel kepada pengguna. Walaupun pengkomputeran awan dikenali sebagai pembekal storan yang utuh bagi pelbagai data, isu utama yang dihadapi adalah cabaran dalam menampung permintaan perkhidmatan dengan ketersediaan data yang tinggi dan akses data yang lebih cepat tanpa mengabaikan sensitiviti data yang disimpan dalam storan awan. Isu-isu ini adalah kritikal kerana ketersediaan dan kebolehcapaian data yang lemah sering mempengaruhi kelewatan capaian data yang akhirnya menjejaskan prestasi sistem.

Bagi membendung isu-isu berbangkit tersebut, hamper semua perkihidmatan awam mengadaptasi strategi 'replikasi data awan' untuk meningkatkan prestasi capaian data yang lebih baik dan menjanjikan kesinambungan perniagaan. Replikasi awan diiktiraf sebagai satu kaedah penyimpanan lebih daripada satu salinan data dalam pelbagai nod storan teragih. Di samping itu, replikasi awan terkenal sebagai teknik komprehensif yang berkeupayaan dalam ketersediaan data yang tinggi, masa capaian yang lebih cepat, toleransi ralat yang lebih baik dan menjanjikan kos yang efektif untuk pengguna dan pembekal awan. Namun demikian, strategi replikasi ini kadangkala mempunyai kecenderungan menjejaskakan prestasi pengkomputeran awan.

Terdapat banyak kajian terdahulu yang menjelaskan isu-isu dalam replikasi awan seperti; kelewatan masa capaian data, masa replikasi yang tinggi, penggunaan ruang storan yang banyak dan penggunaan rangkaian yang tinggi disebabkan pelaksanaan strategi replikasi

yang tidak cekap dalam persekitaran awan. Oleh itu, bagi mengelakkan kebarangkalian negatif dan memenuhi kepuasan hati pengguna, pembangunan strategi replikasi data yang mantap dan efisen adalah signifikan untuk digunakan dalam persekitaran awan bagi meningkatkan prestasi replikasi secara holistik.

Dalam penyelidikan ini, terdapat tiga (3) sumbangan utama dicadangkan untuk persekitaran replikasi awan. Pertama, penyelidikan ini mencadangkan Strategi Replikasi Dipertingkatkan (ERS) yang mempercepatkan masa apabila muat turun fail dalam persekitaran replikasi awan tanpa mengabaikan pemilihan fail yang popular. Kedua, penyelidikan ini mencadangkan teknik penempatan replika yang dinamik Quorum Seimbang (BQ) untuk mengurangkan penggunaan storan tanpa mengabaikan masa replikasi yang lebih cepat dan mengekalkan ketersediaan data secara optimum. Bagi sumbangan ketiga, Kaedah Pemilihan Pusat Data (DCSM) yang cekap dicadangkan untuk memastikan fail sentiasa disimpan di pusat data tempatan yang terbaik. Dengan itu penyelidikan ini dapat mencapai penggunaan rangkaian yang efektif dan meminimakan kekerapan replikasi dalam persekitaran replikasi awan.

ERS, BQ dan DCSM yang dicadangkan adalah komprehensif bergantung kepada keperluan persekitaran replikasi dan menyesuaikan corak akses pengguna secara dinamik yang menjamin data penting sentiasa tersedia dan boleh diakses oleh pengguna. Eksperimen menyeluruh dijalankan menggunakan aplikasi simulasi 'CloudSim'. Keputusan simulasi dianalisa dan graf analitik dibentangkan dalam perbincangan. Untuk mengesahkan kecekapan dan keberkesanan algoritma yang dicadangkan ini, keputusan dibandingkan dengan satu kerja penyelidikan lain dalam bidang yang sama dikenali sebagai algoritma DPRS. Ternyata ERS, BQ dan DCSM yang dicadangkan dalam penyelidikan ini adalah lebih efisen daripada DPRS, dan ianya telah berjaya dibuktikan dengan keputusan masa capaian data yang lebih baik, pengurangan penggunaan storan, masa replikasi yang cepat, mengekalkan ketersediaan data yang tinggi, penggunaan rangkaian yang efektif dan akhirnya dapat meminimumkan kekerapan replikasi dalam persekitaran replikasi awan.

ACKNOWLEDGEMENTS

Alhamdulillah. From the depths of my heart, I thank Allah the Al-Mighty for being my anchor. His omnipresence, faithfulness, and wisdom enabled me to complete my PhD and this thesis.

My heartfelt thanks and appreciation to my supervisors: PM Dr. Rohaya Latip: thank you for the support, guidance, and self-belief you have given me during my difficult times throughout my PhD journey. To Prof. Dr. Hamidah Ibrahim and PM. Dr. Azizol Abdullah: thanks for your valuable views to look at problems from a different light as well as constructive comments on my work. I am not forgetting to thank Dr. Mohamed AlHadi Mahmoud Alrshah, who has been my tutor and guide me on writing good journal papers throughout this PhD journey- May Allah expedite every attempt you make to succeed in life.

I remember with deep appreciation the wind beneath my wings: my husband, Syed Asbar, and my pride and joy, Syed Arham and Syed Arfan, who believed in me, and loved me and all my idiosyncrasies, especially during my most stressful days, my parents: Hj. Mohd Ali & Hjh. Kathoon, my siblings: Bazir, Suaraini, Latifah & Sabrina, my mother-in-law: Hjh. Salmath and in-law family members, my family in Bahau whose prayers and encouragement were always with me.

Special mention goes to Kak Fara, Dalila, Niza, Nisyak, Yatt, Ikin: my friends, companions, motivators, and handy assistants who gave me all the support during my critical encounters -May Allah bless you, girls! Not forgetting other friends/acquaintances who directly or indirectly involved in my PhD journey; Thank you. Last but not least, my closest PhD buddy, Pije whom I could always turn to during my difficult times and rely on for technical assistance throughout my PhD journey. May Allah bless you always.

This thesis was submitted to the Senate of the Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Rohaya binti Latip, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Hamidah binti Ibrahim, PhD
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Azizol bin Abdullah, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

____
ZALILAH MOHD SHARIFF, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 9 June 2022

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software

Signature: ____________________________ Date: _______________

Name and Matric No: Fazlina binti Mohd Ali, GS49592

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: ______________________________

Name of Chairman of Supervisory Committee: Associate Professor Dr. Rohaya binti Latip

Signature: ______________________________

Name of Member of Supervisory Committee: Professor Dr. Hamidah binti Ibrahim

Signature: ______________________________

Name of Member of Supervisory Committee: Associate Professor Dr. Azizol bin Abdullah

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AI	Artificial Intelligent
ART	Average Response Time
BQ	Balance Quorum
BQ	Balance Quorum
CPU	Central Processor Unit
DC	Data Center
DCSM	Data Center Selection Method
DES	Discrete Event Simulation
DFS	Distributed File Systems
DMS	Database Management Systems
ENU	Effective Network Usage
ERS	Enhanced Replication Strategy
FRM	File Request Map
GB	Giga Byte
Gbps	Giga Byte per second
GFS	Google File System
GRM	Global Replica Manager
IaaS	Infrastructure as a Service
IBM	International Business Machines Corporation
ID	Identification Detail
IoT	Internet of Things
IR	Industry Revolution
IT	Information Technology

LRM	Local Replica Manager
LTO	Line Tape-Open
MB	Mega Byte
MCC	Mobile Cloud Computing
P2P	Peer-To-Peer
PaaS	Platform as a Service
PF	Popular Files
QoS	Quality of Service
RAM	Random Access Memory
RF	Replication Frequency
SaaS	Software as a Service
SLA	Service Level Agreement
SLO	Service Level Objective
SU	Storage Usage
VM	Virtual Machines
VPN	Virtual Private Network

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter introduces the concept of cloud computing and data replication strategies. The background of research in current trends is presented. Additionally, the research problem is listed, and objectives for this research work are further explained. The research questions and overall research scope is described for this research. Espectation

1.2 Background

A massive amount of data is used extremely in the current era around the globe. Business researchers like International Data Corporation (IDC) reported that data growth is overwhelming and potentially reaches 35ZB by 2020 (Kaur, Chana, & Bhattacharya, 2018; Nannai John & Mirnalinee, 2020). This phenomenon is derived from the most interconnected Internet of Things (IoT) devices, leading to 100 billion terminals connected in 2025 (Shao, Li, Fu, Jia, & Luo, 2019). The enormous data consist of heterogenous data types from various sources, which are likely recognized as big data (Rajabion, Shaltooki, Taghikhah, Ghasemi, & Badfar, 2019; C. Yang, Huang, Li, Liu, & Hu, 2017). Countless studies embarked in big data emphasised the unsolvable challenge encountered along the decades is focused on data availability and accessibility (Cai, Xu, Jiang, & Vasilakos, 2016; Kaseb, Khafagy, Ali, & Saad, 2019; Slimani, Hamrouni, & Ben Charrada, 2021; Yan, Zhang, Ding, & Zheng, 2017).

Cloud computing is well known as a data management platform to address high-volume data demanded to be accessible by users anytime from anywhere (Mrozek, 2020; Shakarami, Ali, Mohammad, & Hamid, 2021). Cloud computing offers users a "pay-as-you-go" pricing model since it enables multiple services such as Software as a Service (SaaS), which provides real-time application services, Platform as a Service (PaaS), which delivers various operating systems, Consistency as a Service (CaaS), which promises data consistency in storage nodes, and Infrastructure as a Service (IaaS), which provides many hardware solutions to users as an on-demand basis (C. Li, Wang, Tang, & Luo, 2019; Logesswari, Jayanthi, Kalaiselvi, Muthusundari, & Aswin, 2020; N. Mansouri, Javidi, & Mohammad Hasani Zade, 2019; Ostermann & Prodan, 2012; Vikhe & Malhotra, 2015). Cloud computing empowers users with various resilient services that stands vibrantly as preferable technology for almost everyone to ensure data are efficiently managed and business continuity is guaranteed too (Q. Liu, Wang, & Wu, 2014; Nivetha & Vijayakumar, 2016; Tan, Hijazi, Lim, & Gani, 2018). However, cloud computing as a reliable multiple service provider is not exceptional in facing issues in providing high data availability to users while preserving data sensitivity. In fact, fear of losing data during node failures is one of the core issues too (Logesswari et al., 2020; Nannai John & Mirnalinee, 2020; J. Wang, Wu, & Wang, 2017).

Hence, to mitigate the arising concerns in a cloud platform, data replication is recognised as a promising cloud environment strategy (Djebbara & Belbachir, 2018b; Shakarami et al., 2021; Shao et al., 2019; Xie, Yan, & Shen, 2017). Data replication is an empirical technique to accelerate system performance by generating identical data copies across multiple storages (Najme Mansouri, Javidi, & Zade, 2021; Spinnewyn, Botero, & Latre, 2018). Precisely, in a cloud environment, data replication is defined as creating several physical copies for every logical data item and locates the replica copies in different sites or storage nodes (Alami Milani & Jafari Navimipour, 2016; Khelifa, Hamrouni, Mokadem, & Charrada, 2020). There are several ways to apply the data replication function depending on the cloud replication objectives. Respective objectives have drawbacks that occasionally tend to breed performance degradation (Ciritoglu, Saber, Buda, Murphy, & Thorpe, 2018a; N. Mansouri, Javidi, & Mohammad Hasani Zade, 2020; Slimani et al., 2021). Therefore, established and comprehensive replication strategies must be developed to resolve such performance hitches. Literally, a competent replication strategy will allow cloud providers to serve accelerated performance to users with better data availability, faster response time, low fault tolerance, reduced storage usage, and efficient network usage (Alami Milani & Jafari Navimipour, 2017; Fu, Li, Liu, Deng, & Wang, 2019; K. Liu et al., 2020; Malik et al., 2016).

1.3 Problem Statement

Since the past decades, the data replication trend is yet to subside, instead it is progressing rapidly in multiple perspectives to enhance cloud replication performance (Castro-Medina et al., 2019; Seguela, Mokadem, & Pierson, 2020; Slimani et al., 2021). Abundant literature insights show data replication strategies are popular and widespread, leading to fulfilling the demands of cloud replication. The implementation of data replication usually achieves speedy data access, high data availability, efficient network usage, fast replication process, low storage consumptions, and certainly cost-effective in maintaining the entire cloud infrastructure (Fu et al., 2019; J. Liu et al., 2020; Mousavi Nik, Naghibzadeh, & Sedaghat, 2021; Nannai John & Mirnalinee, 2020; Shao et al., 2019; D. W. Sun, Chang, Gao, Jin, & Wang, 2012; Tos, Mokadem, Hameurlain, & Ayav, 2021). Researchers are eagerly focusing on enhancing the strategies in various perceptions; unfortunately, the vulnerability in every strategy is inevitable (P. J. Kumar & Ilango, 2017; N. Mansouri, Javidi, et al., 2020; Slimani et al., 2021; Yi, Wei, & Song, 2017; Zeng & Veeravalli, 2014; Zhang et al., 2021). A non-comprehensive replica strategy would be the main cause to the vulnerability and drawbacks. The drawbacks that usually reside in the developed strategies are not limited to high network usage, high process time, high response time, high storage consumptions and more, depending on the research areas (Alami Milani & Jafari Navimipour, 2016; Runhui Li, Hu, & Lee, 2015; Nannai John & Mirnalinee, 2020; Shao et al., 2019; Tos et al., 2021).

Therefore, the primary issue emphasised in this research is inefficient response time while obtaining data from a cloud replication environment. A good replication strategy must keep substantial data instead of copying the entire data set, leading to replication performance degradation (Karuppusamy & Muthaiyan, 2016; Nannai John & Mirnalinee, 2020; Souravlas & Sifaleras, 2017b). Similarly, many researchers overlooked their strategies while determining crucial data derived explicitly from inefficient response time

(K. A. Kumar, Quamar, Deshpande, & Khuller, 2014; C. Li, Bai, Chen, & Luo, 2020; Long, Zhao, & Chen, 2014; N. Mansouri & Javidi, 2018b; N. Mansouri, Rafsanjani, & Javidi, 2017; Najme Mansouri et al., 2021; Mousavi Nik et al., 2021). Computation overheads commonly trigger the delay in data retrieval during the selection of necessary data files as replica candidates. Nevertheless, some existing research works did not considers substantial factors in determining popular files. As a result, the poor popular file selection strategies affect file download and lead to ineffective response time (Abbes, Louati, & Cérin, 2020; Hussein & Mousa, 2012; Luiz André Barroso, Clidaras, & Urs Hölzle, 2013; N. Mansouri, Javidi, et al., 2020; N. Mansouri et al., 2017; Najme Mansouri, 2016; Najme Mansouri et al., 2021; Shakarami et al., 2021; Shao et al., 2019). Similarly, (N. Mansouri et al., 2017) considered many insignificant factors, such as file size and the total number of jobs requested in every round, which delays their computations. These computations contribute to several additional processes and lead to a higher response time. Furthermore, DPRS includes constant criteria for computations, especially to retrieve the average popularity value in all clusters to find relationships for popular file selection. This constant criterion causes irrelevant processing operations that increase the response time too. The elevated process time eventually defeats the purpose of the replication to accelerate data access by cloud replication users. Therefore, this research work proposed an algorithm to achieve a faster response time with comprehensive popular file selection factors capable of selecting crucial data according to users' access patterns.

The second crucial issue raised in this research is focused on data placement techniques consuming extensive space and high replication time during the placement of replica copies in every storage node in the same cluster environment to achieve high data availability (Kirubakaran, Valarmathy, & Kamalanathan, 2013; N. Mansouri, Mohammad Hasani Zade, & Javidi, 2020; N. Mansouri et al., 2017; Tos et al., 2021). As cloud users always demand high data availability, most of the researchers focused on producing data placement techniques with high numbers of replica copies in almost every cluster and storage nodes (Fu et al., 2019; Gill & Singh, 2016; N. Mansouri et al., 2019; Mousavi Nik et al., 2021; Zeng & Veeravalli, 2014). Eventually, with the developed data placement techniques, most researchers either attained high data availability but, on the other hand, neglected extensive storage usage or vice-versa (C. Li et al., 2020; N. Mansouri et al., 2017; Slimani et al., 2021). Furthermore, in research by (N. Mansouri et al., 2017), the increasing number of replications activities to place replica copies leads to additional replication time too. Therefore, a comprehensive placement technique is required to allocate an adequate number of replica copies while ensuring no extra process time for the replication process. Hence, this research proposed a placement technique that addresses all the raised issues and can reduce storage usage, faster replication time, minimal number of replicas, yet attain high data reliability.

The third problem addressed in this research work is massive network usage due to many data movements in networks to retrieve the requested data by users. The problem was identified when choosing an appropriate data center to store replica copies which were caused by ineffective data center selection criteria (Djebbara & Belbachir, 2018b; Khalajzadeh, Yuan, Zhou, Grundy, & Yang, 2020; N. Mansouri et al., 2017; Najme Mansouri et al., 2021). Researchers proposed many data center criteria to ensure faster access of replica copies in cloud environments. They focused on several substantial factors such as shortest distance or centrality, most requested files, storage availability, and site failures probabilities (Karuppusamy & Muthaiyan, 2016; N. Mansouri et al., 2019, 2017; Nannai John & Mirnalinee, 2020; Tos et al., 2021; Zhang et al., 2021). However, the designated factors in those produced algorithms are inadequate to minimise data movement in the network and preserve fault tolerance. Reducing fault tolerance and data movement between data centers are prevalent challenges in cloud replication environments. These two performances are crucial to mitigate the risk of data loss in networks. Therefore, this research is aware that there is still room for improvements and introduced suitable methods to reduce network usage and decrease replication frequency in existing strategies.

An ineffective replica strategy would result in various drawbacks, as discussed in previous paragraphs. Besides that, it is very decisive to determine how to replicate, how many to replicate, where to replicate, and when to place the replicas in the cloud to ensure the overall replication performance is exclusively attained (Fazlina, Latip, Ibrahim, & Abdullah, 2018; Karunakaran & Kalyanasaravanan, 2018). Hence, this research work proposed solutions for every raised issue through introducing the Enhanced Replication Strategy (ERS) with Balance Quorum (BQ) Replica Placement Technique and Data Center Selection Method (DCSM) for Cloud Replication Environment.

1.4 Research Objectives

The main objective of this thesis is to develop a replication strategy in a Cloud Replication environment that fulfills the following purposes:

1. to propose a replication algorithm that improves response time for file download and appropriate popular file selection.
2. to propose dynamic replica placement techniques to reduce storage usage and response time for file replication and sustain data availability.
3. to propose an efficient data center selection method to achieve efficient network usage and minimise replication frequency, ensuring files are available in the best local data center.

1.5 Research Question

1. What is the best strategy to improve download response time?
2. What is the best technique for replica placement to achieve low storage consumption?
3. What is the suitable method in determining the best data center to store replica copies?

1.6 Research Scope

Numerous latest research works embarked on developing cloud replication strategies by adapting the most advanced techniques such as data mining, fuzzy inference and metaheuristic. Although these revolutionary techniques are recognised as futuristic features, the methods impose a high computation overhead that requires additional computing resources and is known as cost-expensive. Therefore, those distinct techniques were not comprised in this research work.

This thesis focused on a standard data replication strategy in a cloud environment. Furthermore, the critical areas encompassed in this thesis are limited to three (3) primary extents: replication file selection, data placement, and finally replication phase. A detailed explanation of the three (3) main modules/phases are shared in Figure 2.4, Chapter 2. A few areas in the replication strategies were not enclosed in this research, such as data replications that are restricted for download only, thus uploading updates/changes are not considered. Besides that, data deletion and replacement during deficiency of storage were not included in this work. Lastly, data consistency and data security were not within this research scope as well. This research adapted a similar system architecture as (N. Mansouri et al., 2019, 2017; N. Mansouri, Javidi, et al., 2020; Mokadem & Hameurlain, 2020), and the system requirements and parameters were implemented identical to the base work study (N. Mansouri et al., 2017) throughout this research experiment. The details on the system architecture and parameters are shared in Subsections 3.3.3 and 3.3.4, Chapter 3.

Unlike the advanced techniques from other researchers, this research does not require additional costing nor resources. An extensive experiment was conducted for individual contributions in this research. The results showed advancement in response time, storage usage, replication time, data availability, network usage, and replication frequency. Similar to the base work research, simulation was led using the CloudSim simulation tool, and all related performance metrics were proven analytically.

1.7 Thesis Structure

This thesis consists of seven (7) chapters. Chapter One (1) introduces the research background, concepts of cloud computing, and data replication. This chapter is also inclusive of research problems, objectives, research questions, and research scope.

Chapter 2: Literature Review

Chapter Two (2) presents a thorough discussion on the related works that share cloud computing and data replication in a cloud replication environment. Similar research works are further detailed on the proposed algorithms, advantages and disadvantages of the algorithms, and measured performance metrics. Furthermore, data replication types, static and dynamic, are also explained. A complete structure in data replication is enlightened, which comprises selecting popular data, data placement techniques, and data center selection criteria.

Chapter 3: Research Methodology

Chapter Three (3) introduces the methodology used for this research work. The Discrete Event Simulation (DES) lifecycle was adapted as methodology, and the CloudSim simulation tool was selected as the simulation tool for this research. A benchmark study is explained to validate that the simulation environment is identical prior running to the improvement algorithm. The research framework is presented for a holistic understanding of this research contribution. Input/output variables, assumptions, and parameters used in the simulation are listed, and system architecture is illustrated in this chapter too. Finally, related results and graphs are demonstrated for benchmark studies as verification proves.

Chapter 4: Enhance Replication Strategy (ERS)

In Chapter Four (4), the first contribution introduced a replication strategy known as Enhance Replication Strategy (ERS). In this chapter, a detailed ERS flow is explained in phases, and related formulations are introduced. Process diagrams are illustrated for a better understanding of the proposed strategy. A complete algorithm and flowchart are also demonstrated for ERS in this chapter. Detailed experiment results are discussed, and necessary graphs are presented for analytical view and proves.

Chapter 5: Balance Quorum (BQ) Technique

Chapter Five (5) presents the proposed Balance Quorum (BQ) Technique. This second contribution in this research is further discussed, and the related formulations are explained in detail. A process diagram, complete algorithm, and flow chart are presented to understand the BQ technique better. Experiment results and graphs as evidence of the algorithm's success, which outperforms the other research work, are discussed in detail at the end of this chapter.

Chapter 6: Data Center Selection Method (DCSM)

Chapter Six (6) discusses the third proposed contribution known as Data Center Selection Method (DCSM). DCSM comprises three (3) major factors, and the respective factors are explained with structured phases. Every factor inclusive of specific formulations and calculations are illustrated in the process diagram with examples. The complete algorithm and flowchart are presented as a reference in this chapter. Finally, the experiment results are demonstrated in graphs, and the analysis is discussed.

Chapter 7: Conclusion and Future Works

Contributions and conclusions of the overall research are explained in Chapter Seven (7). Then, recommendations on future works are shared as references for upcoming researchers to improvise this similar research area.

REFERENCES

Abbes, H., Louati, T., & Cérin, C. (2020). Dynamic replication factor model for Linux containers-based cloud systems. *Journal of Supercomputing*, *76*(9), 7219–7241. https://doi.org/10.1007/s11227-020-03158-5

Abdulhamid, S. M., & Latiff, M. S. A. (2017). A checkpointed league championship algorithm-based cloud scheduling scheme with secure fault tolerance responsiveness. *Applied Soft Computing Journal*, *61*, 670–680. https://doi.org/10.1016/j.asoc.2017.08.048

Al-Dhuraibi, Y., Paraiso, F., Djarallah, N., & Merle, P. (2018). Elasticity in Cloud Computing: State of the Art and Research Challenges. *IEEE Transactions on Services Computing*. https://doi.org/10.1109/TSC.2017.2711009

Alami Milani, B., & Jafari Navimipour, N. (2016). A comprehensive review of the data replication techniques in the cloud environments: Major trends and future directions. *Journal of Network and Computer Applications*, *64*, 229–238. https://doi.org/10.1016/j.jnca.2016.02.005

Alami Milani, B., & Jafari Navimipour, N. (2017). A Systematic Literature Review of the Data Replication Techniques in the Cloud Environments. *Big Data Research*, *10*(C), 1–7. https://doi.org/10.1016/j.bdr.2017.06.003

Alli, A. A., & Alam, M. M. (2020). The fog cloud of things: A survey on concepts, architecture, standards, tools, and applications. *Internet of Things*, *9*, 100177. https://doi.org/10.1016/j.iot.2020.100177

Alshammari, M. M., Alwan, A. A., Nordin, A., & Abualkishik, A. Z. (2018). Disaster Recovery with Minimum Replica Plan for Reliability Checking in Multi-Cloud. *Procedia Computer Science*, *130*, 247–254. https://doi.org/10.1016/j.procs.2018.04.036

Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A. S., & Buyya, R. (2015). Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, *79–80*, 3–15. https://doi.org/10.1016/j.jpdc.2014.08.003

Atrey, A., Van Seghbroeck, G., Mora, H., Volckaert, B., & De Turck, F. (2020). UnifyDR: A Generic Framework for Unifying Data and Replica Placement. *IEEE Access*, *8*(1), 216894–216910. https://doi.org/10.1109/ACCESS.2020.3041670

Awad, A., Salem, R., Abdelkader, H., & Salam, M. A. (2021). A Novel Intelligent Approach for Dynamic Data Replication in Cloud Environment. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2021.3064917

Boru, D., Kliazovich, D., Granelli, F., Bouvry, P., & Zomaya, A. Y. (2015). Energy-efficient data replication in cloud computing datacenters. *Cluster Computing*, *18*(1), 385–402. https://doi.org/10.1007/s10586-014-0404-x

Bowers, A., Liao, C., Steiert, D., Lin, D., Squicciarini, A., & Hurson, A. (2021). Detecting Suspicious File Migration or Replication in the Cloud. *IEEE Transaction on Dependable and Secure Computing*, *Vol. 18*, *N*(1), 296–309.

Breslau, L., Cao, P., Fan, L., Phillips, G., & Shenker, S. (1999). Web caching and zipf-like distributions: Evidence and implications. *Proceedings - IEEE INFOCOM*, *1*, 126–134. https://doi.org/10.1109/INFCOM.1999.749260

C.Venish Raja, D. L. J. (2020). A Cost Effective Scalable Scheme for Dynamic Data Service in Heterogeneous Cloud Environment. *International Journal of Advanced Science and Technology*, *Vol. 28*, *N*(January).

Cai, H., Xu, B., Jiang, L., & Vasilakos, A. V. (2016). IoT-based Big Data Storage Systems in Cloud Computing: Perspectives and Challenges. *IEEE Internet of Things Journal*, *PP*(99), 1–1. https://doi.org/10.1109/JIOT.2016.2619369

Calheiros, R. N., Ranjan, R., Beloglazov, A., De Rose, C. A. F., & Buyya, R. (2011). CloudSim: A toolkit for modelling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Software - Practice and Experience*. https://doi.org/10.1002/spe.995

Carns, P., Harms, K., Jenkins, J., Mubarak, M., Ross, R., & Carothers, C. (2017). Impact of data placement on resilience in large-scale object storage systems. *IEEE Symposium on Mass Storage Systems and Technologies*. https://doi.org/10.1109/MSST.2016.7897091

Castro-Medina, F., Rodriguez-Mazahua, L., Abud-Figueroa, M. A., Romero-Torres, C., Reyes-Hernandez, L. A., & Alor-Hernandez, G. (2019). Application of data fragmentation and replication methods in the cloud: A review. In *CONIELECOMP 2019 - 2019 International Conference on Electronics, Communications and Computers* (pp. 47–54). https://doi.org/10.1109/CONIELECOMP.2019.8673249

Chang, V. (2015). Towards a Big Data system disaster recovery in a Private Cloud. *Ad Hoc Networks*, *35*, 65–82. https://doi.org/10.1016/j.adhoc.2015.07.012

Chaturvedi, N. (2012). Analysis of Replication and Replication Algorithms in Distributed System, *2*(5), 261–266.

Chen, T., Bahsoon, R., & Tawil, A. R. H. (2014). Scalable service-oriented replication with flexible consistency guarantee in the cloud. *Information Sciences*, *264*, 349–370. https://doi.org/10.1016/j.ins.2013.11.024

Cidon, A., Stutsman, R., Rumble, S., Katti, S., Ousterhout, J., & Rosenblum, M. (2013). MinCopysets : Derandomizing Replication In Cloud Storage. *Networked Systems Design and Impementation (NSDI)*.

Cirani, S., Ferrari, G., Mancin, M., & Picone, M. (2018). Virtual Replication of IoT Hubs in the Cloud: A Flexible Approach to Smart Object Management. *Journal of Sensor and Actuator Networks*, *7*(2), 16. https://doi.org/10.3390/jsan7020016

Ciritoglu, H. E., Saber, T., Buda, T. S., Murphy, J., & Thorpe, C. (2018a). Towards a Better Replica Management for Hadoop Distributed File System. In *in 2018 IEEE International Congress on Big Data (BigData Congress)* (pp. 104–111). https://doi.org/10.1109/BigDataCongress.2018.00021

Ciritoglu, H. E., Saber, T., Buda, T. S., Murphy, J., & Thorpe, C. (2018b). Towards a Better Replica Management for Hadoop Distributed File System. In *Proceedings - 2018 IEEE International Congress on Big Data, BigData Congress 2018 - Part of the 2018 IEEE World Congress on Services* (pp. 104–111). https://doi.org/10.1109/BigDataCongress.2018.00021

Craft, R. C., & Leake, C. (2002). The Pareto principle in organizational decision making. *Management Decision*, *40*(8), 729–733. https://doi.org/10.1108/00251740210437699

Djebbara, M. R., & Belbachir, H. (2018a). Cloud data replication strategy using multiple-criteria decision analysis methods. *Multiagent and Grid Systems*, *14*(2), 203–218. https://doi.org/10.3233/mgs-180288

Djebbara, M. R., & Belbachir, H. (2018b). Cost Function Based On Analytic Hierarchy Process for Data Replication Strategy in Cloud Environment. *Journal of Theoretical and Applied Information Technology*, *96*, 2638–2648.

Ebadi, Y., & Jafari Navimipour, N. (2018). An energy-aware method for data replication in the cloud environments using a Tabu search and particle swarm optimization algorithm. *Concurrency and Computation: Practice and Experience*, (July 2017), e4757. https://doi.org/10.1002/cpe.4757

Elgendy, I. A., Zhang, W. Z., He, H., Gupta, B. B., & Abd El-Latif, A. A. (2021). Joint computation offloading and task caching for multi-user and multi-task MEC systems: reinforcement learning-based algorithms. *Wireless Network*, *27*(3), 2023–2038. https://doi.org/10.1007/s11276-021-02554-w

Fazlina, M. A., Latip, R., Ibrahim, H., & Abdullah, A. (2018). A Review : Replication Strategies for Big Data in Cloud Environment. *International Journal of Engineering & Technology*, *7*, 357–362.

Filho, M. C. S., Oliveira, R. L., & Monteiro, C. C. (2018). CloudSim Plus Documentation Release 1.0.0. Retrieved from https://media.readthedocs.org/pdf/cloudsimplus/latest/cloudsimplus.pdf

Fu, X., Li, J., Liu, W., Deng, S., & Wang, J. (2019). Data Replica Placement Policy Based on Load Balance in Cloud Storage System. *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, (Itnec), 682–685. https://doi.org/10.1109/ITNEC.2019.8728995

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

George, S., & Edwin, E. B. (2018). A Review on Data Replication Strategy in Cloud Computing. *2017 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2017*, 1–4. https://doi.org/10.1109/ICCIC.2017.8524269

Ghemawat, S., Gobioff, H., & Leung, S.-T. (2003). The Google file system. *Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles - SOSP '03*, 29. https://doi.org/10.1145/945449.945450

Gholipour, N., Arianyan, E., & Buyya, R. (2020). A novel energy-aware resource management technique using joint VM and container consolidation approach for green computing in cloud data centers. *Simulation Modelling Practice and Theory*, *104*(March), 102127. https://doi.org/10.1016/j.simpat.2020.102127

Gill, N. K., & Singh, S. (2016). A dynamic, cost-aware, optimized data replication strategy for heterogeneous cloud data centers. *Future Generation Computer Systems*, *65*, 10–32. https://doi.org/10.1016/j.future.2016.05.016

Gode, R. V, & Dalvi, R. (2017). An effective storage management in a twin cloud architecture using an authorized deduplication technique. *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, 796–801.

Gopinath, S., & Sherly, E. (2017). A weighted dynamic data replication management for cloud data storage systems. *International Journal of Applied Engineering Research*, *12*(24), 15517–15524.

Gopinath, S., & Sherly, E. (2018). ScienceDirect ScienceDirect International Conference on Computational Intelligence and Data Science (ICCIDS 2018) A Dynamic Replica Factor Calculator for Weighted Dynamic A Dynamic Replica Factor Calculator for Weighted Dynamic Replication Management in. *Procedia Computer Science*, *132*, 1771–1780. https://doi.org/10.1016/j.procs.2018.05.152

Gorbenko, A., Romanovsky, A., & Tarasyuk, O. (2019). Fault tolerant internet computing: Benchmarking and modelling trade-offs between availability, latency and consistency. *Journal of Network and Computer Applications*, *146*(July), 102412. https://doi.org/10.1016/j.jnca.2019.102412

Gracia-Tinedo, R., Artigas, M. S., & García López, P. (2012). Analysis of data availability in F2F storage systems: When correlations matter. *2012 IEEE 12th International Conference on Peer-to-Peer Computing, P2P 2012*, (September), 225–236. https://doi.org/10.1109/P2P.2012.6335803

Grolinger, K., Higashino, W. A., Tiwari, A., & Capretz, M. A. M. (2013). Data management in cloud environments: NoSQL and NewSQL data stores. *Journal of Cloud Computing*, *2*(1). https://doi.org/10.1186/2192-113X-2-22

Habibi, F., & Kumar, N. (2017). A brief survey on dynamic strategies of data replication in cloud environment : last five year study. *International Journal of Engineering Development and Research*, *5*(4), 342–345.

Hage, P., & Harary, F. (1995). Eccentricity and centrality in networks. *Social Networks*, *17*(1), 57–63. https://doi.org/10.1016/0378-8733(94)00248-9

Hanandeh, F., Khazaaleh, M., Ibrahim, H., & Latip, R. (2012). CFS: A new dynamic replication strategy for data grids. *International Arab Journal of Information Technology*, *9*(1).

Hasan, M., & Goraya, M. S. (2018). Fault tolerance in cloud computing environment: A systematic survey. *Computers in Industry*, *99*(March), 156–172. https://doi.org/10.1016/j.compind.2018.03.027

He, L., Qian, Z., & Shang, F. (2018). A novel predicted replication strategy in cloud storage. *Journal of Supercomputing*, *76*(7), 4838–4856. https://doi.org/10.1007/s11227-018-2647-4

Huang, C., Chen, W., Yuan, L., Ding, Y., Jian, S., Tan, Y., … Chen, D. (2020). Toward security as a service: A trusted cloud service architecture with policy customization. *Journal of Parallel and Distributed Computing*, *Volume 149*, Pages 76-88. https://doi.org/10.1016/j.jpdc.2020.11.002

Hussein, M., & Mousa, M. (2012). A Light-weight Data Replication for Cloud Data Centers Environment. *International Journal of Engineering and Innovative Technology (IJEIT)*, *1*(6), 169–175. Retrieved from http://ijeit.com/vol 1/Issue 6/IJEIT1412201206_31.pdf

Ibrahim, F. A. M., & Hemayed, E. E. (2019). Trusted Cloud Computing Architectures for infrastructure as a service: Survey and systematic literature review. *Computers and Security*, *82*, 196–226. https://doi.org/10.1016/j.cose.2018.12.014

J. Sridharan, C. Valliyammai, R. N. K. and L. N. K. (2019). Data De-duplication Using Cuckoo Hashing in Cloud Storage. *Soft Computing in Data Analytics*, *758*, Soft Comput. Data Anal. https://doi.org/10.1007/978-981-13-0514-6

Jayapandian, N., Rahman, A. M. J. M. Z., & Nandhini, I. (2016). A novel approach for handling sensitive data with deduplication method in hybrid cloud. *IC-GET 2015 - Proceedings of 2015 Online International Conference on Green Engineering and Technologies*, 1–6. https://doi.org/10.1109/GET.2015.7453774

Ji, C., Li, Y., Qiu, W., Awada, U., & Li, K. (2012). Big data processing in cloud computing environments. *Proceedings of the 2012 International Symposium on Pervasive Systems, Algorithms, and Networks, I-SPAN 2012*, 17–23. https://doi.org/10.1109/I-SPAN.2012.9

Jong, J. J., Park, H., Arabnia, H. R., & Hutchison, D. (2013). Grid and Pervasive Computing. In *8th International Conference, GPC 2013 and Colocated Workshops, Seoul, Korea, Proceedings*. Springer, Berlin, Heidelberg.

Julisch, K., & Hall, M. (2010). Security and control in the cloud. *Information Security Journal*, *19*(6), 299–309. https://doi.org/10.1080/19393555.2010.514654

Jun, Y., & Lihong, Y. (2017). The Cloud Technology Double Live Data Center Information System Research and Design Based on Disaster Recovery Platform. *Procedia Engineering*, *174*, 1356–1370. https://doi.org/10.1016/j.proeng.2017.01.289

Karunakaran, S., & Kalyanasaravanan, A. (2018). Active Data Replication in Cloud to Enhance the Availability of Popular Facts. *Journal of Adroid and IOS Applications and Testing*, *3*(1), 1–8.

Karuppusamy, S., & Muthaiyan, M. (2016). An efficient placement algorithm for data replication and to improve system availability in cloud environment. *International Journal of Intelligent Engineering and Systems*, *9*(4), 88–97. https://doi.org/10.22266/ijies2016.1231.10

Kaseb, M. R., Khafagy, M. H., Ali, I. A., & Saad, E. S. M. (2019). An improved technique for increasing availability in Big Data replication. *Future Generation Computer Systems*, *91*, 493–505. https://doi.org/10.1016/j.future.2018.08.015

Kaur, R., Chana, I., & Bhattacharya, J. (2018). Data deduplication techniques for efficient cloud storage management: a systematic review. *Journal of Supercomputing*, *74*(5), 2035–2085. https://doi.org/10.1007/s11227-017-2210-8

Khalajzadeh, H., Yuan, D., Zhou, B. B., Grundy, J., & Yang, Y. (2020). Cost effective dynamic data placement for efficient access of social networks. *Journal of Parallel and Distributed Computing*, *141*, 82–98. https://doi.org/10.1016/j.jpdc.2020.03.013

Khan, M., Wu, X., Xu, X., & Dou, W. (2017). Big data challenges and opportunities in the hype of Industry 4.0. *2017 IEEE International Conference on Communications (ICC)*, 1–6. https://doi.org/10.1109/ICC.2017.7996801

Khelifa, A., Hamrouni, T., Mokadem, R., & Charrada, F. Ben. (2020). SLA-aware task scheduling and data replication for enhancing provider profit in clouds. *Procedia Computer Science*, *176*, 3143–3152. https://doi.org/10.1016/j.procs.2020.09.174

Kirubakaran, S., Valarmathy, S., & Kamalanathan, C. (2013). Data replication using modified D2RS in cloud computing for performance improvement. *Journal of Theoretical and Applied Information Technology*.

Kumar, K. A., Quamar, A., Deshpande, A., & Khuller, S. (2014). SWORD: workload-aware data placement and replica selection for cloud data management systems. *VLDB Journal*, *23*(6), 845–870. https://doi.org/10.1007/s00778-014-0362-1

Kumar, P. J., & Ilango, P. (2017). Data Replication in Current Generation Computing Environment. *International Journal of Engineering Trends and Technology*, *45*(10), 488–492. https://doi.org/10.14445/22315381/ijett-v45p292

Laouar, K. T. R. M. M. R. (2019). A Data Replication Strategy for Document Oriented NoSQL Databases. *International Journal of Grid and Utility Computing*, *10*(1), 36–41. https://doi.org/10.1504/IJGUC.2019.097224

Latip, R., Mizan, T., Ghazali, N. F., Abd Kadir, R., & Hanandeh, F. A. (2013). Replica control protocol: Triple Quorum Replication (TQR) in Data Grid. *2013 5th International Conference on Computer Science and Information Technology, CSIT 2013 - Proceedings*, 303–307. https://doi.org/10.1109/CSIT.2013.6588796

Latip, R., Othman, M., Abdullah, A., Ibrahim, H., & Sulaiman, M. N. (2009). Quorum-based data replication in grid environment. *International Journal of Computational Intelligence Systems*, *2*(4), 386–397. https://doi.org/10.1080/18756891.2009.9727670

Lauer, M. (2011). Data Security in the Cloud Why Cloud Computing ? *Seminar*, 61–64. https://doi.org/10.1109/MSP.2009.87

Lei, M., Vrbsky, S. V., & Hong, X. (2008). An on-line replication strategy to increase availability in Data Grids. *Future Generation Computer Systems*, *24*(2), 85–98. https://doi.org/10.1016/j.future.2007.04.009

Levitin, G., Xing, L., & Luo, L. (2019). Joint optimal checkpointing and rejuvenation policy for real-time computing tasks. *Reliability Engineering and System Safety*, *182*(October 2018), 63–72. https://doi.org/10.1016/j.ress.2018.10.006

Li, C., Bai, J., Chen, Y., & Luo, Y. (2020). Resource and replica management strategy for optimizing financial cost and user experience in edge cloud computing system. *Information Sciences*, *516*, 33–55. https://doi.org/10.1016/j.ins.2019.12.049

Li, C., Tang, J., & Luo, Y. (2019). Scalable replica selection based on node service capability for improving data access performance in edge computing environment. *The Journal of Supercomputing, 75*, 7209–7243. https://doi.org/10.1007/s11227-019-02930-6

Li, C., Wang, Y., Tang, H., & Luo, Y. (2019). Dynamic multi-objective optimized replica placement and migration strategies for SaaS applications in edge cloud. *Future Generation Computer Systems*, *100*, 921–938. https://doi.org/10.1016/j.future.2019.05.003

Li, C., Wang, Y., Tang, H., Zhang, Y., Xin, Y., & Luo, Y. (2019). Flexible replica placement for enhancing the availability in edge computing environment. *Computer Communications*. https://doi.org/10.1016/j.comcom.2019.07.013

Li, Jianjiang, Zhang, P., Li, Y., Chen, W., Liu, Y., & Wang, L. (2017). A data-check based distributed storage model for storing hot temporary data. *Future Generation Computer Systems*, *73*, 13–21. https://doi.org/10.1016/j.future.2017.03.019

Li, Jin, Li, Y. K., Chen, X., Lee, P. P. C., & Lou, W. (2015). A Hybrid Cloud Approach for Secure Authorized Deduplication. *IEEE Transactions on Parallel and Distributed Systems*, *26*(5), 1206–1216. https://doi.org/10.1109/TPDS.2014.2318320

Li, Rongheng, Zhang, J., & Shen, W. (2018). Replicas Strategy and Cache Optimization of Video Surveillance Systems Based on Cloud Storage. *Future Internet*, *10*(4), 34. https://doi.org/10.3390/fi10040034

Li, Runhui, Hu, Y., & Lee, P. P. C. (2015). Enabling Efficient and Reliable Transition from Replication to Erasure Coding for Clustered File Systems. *2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, *28*(9), 148–159. https://doi.org/10.1109/DSN.2015.24

Li, W., Yang, Y., & Yuan, D. (2016). Ensuring Cloud Data Reliability with Minimum Replication by Proactive Replica Checking. *IEEE Transactions on Computers*, *65*(5), 1494–1506. https://doi.org/10.1109/TC.2015.2451644

Li, X., Jiang, X., Garraghan, P., & Wu, Z. (2018). Holistic energy and failure aware workload scheduling in Cloud datacenters. *Future Generation Computer Systems*, *78*, 887–900. https://doi.org/10.1016/j.future.2017.07.044

Limam, S., Mokadem, R., & Belalem, G. (2019). Data replication strategy with satisfaction of availability, performance and tenant budget requirements. *Cluster Computing*. https://doi.org/10.1007/s10586-018-02899-6

Liu, J., Shen, H., Chi, H., Narman, H. S., Yang, Y., Cheng, L., & Chung, W. (2020). A Low-Cost Multi-Failure Resilient Replication Scheme for High-Data Availability in Cloud Storage. *IEEE/ACM Transactions on Networking*, 1–16. https://doi.org/10.1109/TNET.2020.3027814

Liu, J., Shen, H., & Narman, H. S. (2019). Popularity-Aware Multi-Failure Resilient and Cost-Effective Replication for High Data Durability in Cloud Storage. *IEEE Transactions on Parallel and Distributed Systems*, *30*(10), 2355–2369. https://doi.org/10.1109/TPDS.2018.2873384

Liu, K., Peng, J., Wang, J., Liu, W., Huang, Z., & Pan, J. (2020). Scalable and adaptive data replica placement for geo-distributed cloud storages. *IEEE Transactions on Parallel and Distributed Systems*, *31*(7), 1575–1587. https://doi.org/10.1109/TPDS.2020.2968321

Liu, Q., Wang, G., & Wu, J. (2014). Consistency as a service: Auditing cloud consistency. *IEEE Transactions on Network and Service Management*, *11*(1), 25–35. https://doi.org/10.1109/TNSM.2013.122613.130411

Logesswari, S., Jayanthi, S., Kalaiselvi, D., Muthusundari, S., & Aswin, V. (2020). Materials Today : Proceedings A study on cloud computing challenges and its mitigations. *Materials Today: Proceedings*, (xxxx). https://doi.org/10.1016/j.matpr.2020.10.655

Londhe, A., Bhalerao, V., Ghodey, S., Kate, S., Dandekar, N., & Bhange, S. (2019). Data Division and Replication Approach for Improving Security and Availability of Cloud Storage. In *Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018* (pp. 1–4). IEEE. https://doi.org/10.1109/ICCUBEA.2018.8697355

Long, S. Q., Zhao, Y. L., & Chen, W. (2014). MORM: A Multi-objective Optimized Replication Management strategy for cloud storage cluster. *Journal of Systems Architecture*, *60*(2), 234–244. https://doi.org/10.1016/j.sysarc.2013.11.012

Lu, S., Wu, J., Duan, Y., Wang, N., & Fang, J. (2020). Towards cost-efficient resource provisioning with multiple mobile users in fog computing. *Journal of Parallel and Distributed Computing*, *146*, 96–106. https://doi.org/10.1016/j.jpdc.2020.08.002

Lui, F., Tong, J., Mao, J., Bohn, R., Messina, J., Badger, L., & Leaf, D. (2011). *NIST Cloud Computing Reference Architecture: Recommendations of NIST. National Institute of Standard and Technology* (Vol. NIST Speci).

Luiz André Barroso, Clidaras, J., & Urs Hölzle. (2013). The datacenter as a computer: an introduction to the design of warehouse-scale machines. *Synthesis Lectures on Computer Architecture*, *8*(3), 1–154. https://doi.org/10.2200/S00516ED2V01Y201306CAC024

Lynn, T., Rosati, P., Lejeune, A., & Emeakaroha, V. (2017). A Preliminary Review of Enterprise Serverless Cloud Computing (Function-as-a-Service) Platforms. *Proceedings of the International Conference on Cloud Computing Technology and Science, CloudCom*, *2017-Decem*, 162–169. https://doi.org/10.1109/CloudCom.2017.15

Ma, J., Wang, G., & Liu, X. (2016). DedupeSwift: Object-oriented storage system based on data deduplication. In *IEEE Trustcom/BigDataSE/ISPA* (pp. 1069–1076). IEEE. https://doi.org/10.1109/TrustCom.2016.0177

Malik, S. U. R., Khan, S. U., Ewen, S. J., Tziritas, N., Kolodziej, J., Zomaya, A. Y., … Li, H. (2016). Performance analysis of data intensive cloud systems based on data management and replication: a survey. *Distributed and Parallel Databases*, *34*(2), 179–215. https://doi.org/10.1007/s10619-015-7173-2

Mansouri, N., & Dastghaibyfard, G. (2011). *A New Dynamic Replication Algorithm for Hierarchy Networks in Data Grid*. *2011 International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*. IEEE. https://doi.org/10.1109/3PGCIC.2011.36

Mansouri, N., Ghafari, R., & Zade, B. M. H. (2020). Cloud computing simulators: A comprehensive review. *Simulation Modelling Practice and Theory*, *104*(July). https://doi.org/10.1016/j.simpat.2020.102144

Mansouri, N., & Javidi, M. M. (2018a). A hybrid data replication strategy with fuzzy-based deletion for heterogeneous cloud data centers. *Journal of Supercomputing*, *74*(10), 5349–5372. https://doi.org/10.1007/s11227-018-2427-1

Mansouri, N., & Javidi, M. M. (2018b). A new Prefetching-aware Data Replication to decrease access latency in cloud environment. *The Journal of Systems & Software*. https://doi.org/10.1016/j.jss.2018.05.027

Mansouri, N., Javidi, M. M., & Mohammad Hasani Zade, B. (2019). Using data mining techniques to improve replica management in cloud environment. *Soft Computing*, *0123456789*. https://doi.org/10.1007/s00500-019-04357-w

Mansouri, N., Javidi, M. M., & Mohammad Hasani Zade, B. A. (2020). A CSO‑based approach for secure data replication in cloud. *The Journal of Supercomputing*, *77*(6), 5882–5933. https://doi.org/10.1007/s11227-020-03497-3

Mansouri, N., Mohammad Hasani Zade, B., & Javidi, M. M. (2020). A multi-objective optimized replication using fuzzy based self-defense algorithm for cloud computing. *Journal of Network and Computer Applications*, *171*(October 2019), 102811. https://doi.org/10.1016/j.jnca.2020.102811

Mansouri, N., Rafsanjani, M. K., & Javidi, M. M. (2017). DPRS: A dynamic popularity aware replication strategy with parallel download scheme in cloud environments. *Simulation Modelling Practice and Theory*, *77*, 177–196. https://doi.org/10.1016/j.simpat.2017.06.001

Mansouri, N ., & Javidi, M. M. (2019). Improving Data Availability Using Combined Replication Strategy in Cloud Environment. *Iranian Journal of Electrical and Electronic Engineering*, *15*(3), 282–293. https://doi.org/10.22068/IJEEE.15.3.282

Mansouri, Najme. (2016). Adaptive data replication strategy in cloud computing for performance improvement. *Frontiers of Computer Science*, *10*(5), 925–935. https://doi.org/10.1007/s11704-016-5182-6

Mansouri, Najme, Javidi, M. M., & Zade, B. M. H. (2021). Hierarchical data replication strategy to improve performance in cloud computing. *Frontiers of Computer Science*, *15*(2). https://doi.org/10.1007/s11704-019-9099-8

Maruti, M. V., & Nighot, M. K. (2016). Authorized data Deduplication using hybrid cloud technique. *International Conference on Energy Systems and Applications, ICESA 2015*, (Icesa), 695–699. https://doi.org/10.1109/ICESA.2015.7503439

Mazumdar, S., Seybold, D., Kritikos, K., & Verginadis, Y. (2019). *A survey on data storage and placement methodologies for Cloud-Big Data ecosystem. Journal of Big Data* (Vol. 6). Springer International Publishing. https://doi.org/10.1186/s40537-019-0178-3

Mesbahi, M. R., Rahmani, A. M., & Hosseinzadeh, M. (2017). Highly reliable architecture using the 80/20 rule in cloud computing datacenters. *Future Generation Computer Systems*, *77*, 77–86. https://doi.org/10.1016/j.future.2017.06.011

Miao, Y., Wu, G., Li, M., Ghoneim, A., Al-Rakhami, M., & Hossain, M. S. (2020). Intelligent task prediction and computation offloading based on mobile-edge cloud computing. *Future Generation Computer Systems*, *102*, 925–931. https://doi.org/10.1016/j.future.2019.09.035

Mishra, P., & Somani, A. K. (2017). Host managed contention avoidance storage solutions for Big Data. *Journal of Big Data*, *4*(1). https://doi.org/10.1186/s40537-017-0080-9

Mishra, S. (2017). On using CloudSim as a Cloud Simulator : The Manual, (December). https://doi.org/10.13140/RG.2.2.30215.91041

Mokadem, R., & Hameurlain, A. (2020). A data replication strategy with tenant performance and provider economic profit guarantees in Cloud data centers. *Journal of Systems and Software*, *159*, 110447. https://doi.org/10.1016/j.jss.2019.110447

Mousavi Nik, S. S., Naghibzadeh, M., & Sedaghat, Y. (2021). Task replication to improve the reliability of running workflows on the cloud. *Cluster Computing*, *24*(1), 343–359. https://doi.org/10.1007/s10586-020-03109-y

Mrozek, D. (2020). A review of Cloud computing technologies for comprehensive microRNA analyses. *Computational Biology and Chemistry*, *88*(July), 107365. https://doi.org/10.1016/j.compbiolchem.2020.107365

Nannai John, S., & Mirnalinee, T. T. (2020). A novel dynamic data replication strategy to improve access efficiency of cloud storage. *Information Systems and E-Business Management*, *18*(3), 405–426. https://doi.org/10.1007/s10257-019-00422-x

Newman, M. E. J. (2009). Networks of networks - An introduction. *(2010, Oxford University Press.) Artificial Life.*, *80*, 1–6. https://doi.org/10.1016/j.chaos.2015.03.016

Nirmala, S. J., Setlur, A. R., Singh, H. S., & Khoriya, S. (2020). An efficient fault tolerant workflow scheduling approach using replication heuristics and checkpointing in the cloud. *Journal of Parallel and Distributed Computing*, *136*, 14–28. https://doi.org/10.1016/j.jpdc.2019.09.004

Nivetha, N. K., & Vijayakumar, D. (2016). Modeling fuzzy based replication strategy to improve data availabiity in cloud datacenter. In *2016 International Conference on Computing Technologies and Intelligent Data Engineering, ICCTIDE 2016* (Vol. 1, pp. 1–6). IEEE. https://doi.org/10.1109/ICCTIDE.2016.7725322

O'Donovan, P., Gallagher, C., Leahy, K., & O'Sullivan, D. T. J. (2019). A comparison of fog and cloud computing cyber-physical interfaces for Industry 4.0 real-time embedded machine learning engineering applications. *Computers in Industry*, *110*, 12–35. https://doi.org/10.1016/j.compind.2019.04.016

Ostermann, S., & Prodan, R. (2012). *Impact of variable priced cloud resources on scientific workflow scheduling*. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 7484 LNCS). https://doi.org/10.1007/978-3-642-32820-6_35

Overby, E., Bharadwaj, A., & Sambamurthy, V. (2006). Enterprise agility and the enabling role of information technology. *European Journal of Information Systems*, *15*(2), 120–131. https://doi.org/10.1057/palgrave.ejis.3000600

Pezoulas, V. C., Exarchos, T. P., & Fotiadis, D. I. (2020). *Cloud infrastructures for data sharing*. *Medical Data Sharing, Harmonization and Analytics*. https://doi.org/10.1016/b978-0-12-816507-2.00006-2

Radi, M. (2017). Replica Placement Strategy Based on Analytic Hierarchy Process in Heterogeneous Cloud Data Storage. *International Arab Conference on Information Technology (ACIT ' 2016)*, (December 2016).

Rahimi, M., Songhorabadi, M., & Kashani, M. H. (2020). Fog-based smart homes: A systematic review. *Journal of Network and Computer Applications*, *153*(November 2019), 102531. https://doi.org/10.1016/j.jnca.2020.102531

Rajabion, L., Shaltooki, A. A., Taghikhah, M., Ghasemi, A., & Badfar, A. (2019). Healthcare big data processing mechanisms: The role of cloud computing. *International Journal of Information Management*, *49*(June 2017), 271–289. https://doi.org/10.1016/j.ijinfomgt.2019.05.017

Ramachandran, M., & Chang, V. (2016). Towards performance evaluation of cloud service providers for cloud data security. *International Journal of Information Management*, *36*(4), 618–625. https://doi.org/10.1016/j.ijinfomgt.2016.03.005

Rohani, H., & Roosta, A. K. (2014). Calculating Total System Availability. *Universiteit van Amsterdam*. Retrieved from https://staff.science.uva.nl/c.t.a.m.delaat/rp/2013-2014/p17/report.pdf

Saleh, A., Javidan, R., & FatehiKhajeh, M. T. (2015). A four-phase data replication algorithm for data grid. *Journal of Advanced Computer Science & Technology*, *4*(1), 163. https://doi.org/10.14419/jacst.v4i1.4009

Sargent, R. G. (2010). Verification and validation of simulation models. *Proceedings - Winter Simulation Conference*, (January 2011), 166–183. https://doi.org/10.1109/WSC.2010.5679166

Saritha, K., & Subasree, S. (2015). Analysis of hybrid cloud approach for private cloud in the de-duplication mechanism. *ICETECH 2015 - 2015 IEEE International Conference on Engineering and Technology*, (March), 1–3. https://doi.org/10.1109/ICETECH.2015.7275010

Seguela, M., Mokadem, R., & Pierson, J.-M. (2020). Comparing energy-aware vs. cost-aware data replication strategy. *International Green and Sustainable Computing Conference (IGSC)*, 1–8. https://doi.org/10.1109/igsc48788.2019.8957206

Selvi, S. A. E., & Anbuselvi, R. (2018a). Popularity (Hit Rate) based replica creation for enhancing the availability in cloud storage. *International Journal of Intelligent Engineering and Systems*, *11*(2), 161–172. https://doi.org/10.22266/IJIES2018.0430.18

Selvi, S. A. E., & Anbuselvi, R. (2018b). Popularity (Hit Rate) based replica creation for enhancing the availability in cloud storage. *International Journal of Intelligent Engineering and Systems*, *11*(2), 161–172. https://doi.org/10.22266/IJIES2018.0430.18

Shakarami, A., Ali, M. G., Mohammad, S., & Hamid, M. (2021). Data Replication Schemes in Cloud Computing : A Survey. *Cluster Computing*, *7*. https://doi.org/10.1007/s10586-021-03283-7

Shao, Y., Li, C., Fu, Z., Jia, L., & Luo, Y. (2019). Cost-effective replication management and scheduling in edge computing. *Journal of Network and Computer Applications*, *129*(May 2018), 46–61. https://doi.org/10.1016/j.jnca.2019.01.001

Shorfuzzaman, M. (2017). On the Dynamic Maintenance of Data Replicas based on Access Patterns in A Multi-Cloud Environment. *International Journal of Advanced Computer Science and Applications (IJACSA)*, *8*(3), 207–215.

Shorfuzzaman, M., & Masud, M. (2019). Leveraging a multi-objective approach to data replication in cloud computing environment to support big data applications. *International Journal of Advanced Computer Science and Applications*, *10*(3), 418–429. https://doi.org/10.14569/IJACSA.2019.0100354

Singh, A. (2019). Comparative Study of Concurrency and Replica control protocols in Distributed Environment, *14*(2), 329–334. Retrieved from http://www.ripublication.com

Singh, L., & Malhotra, J. (2016). A Survey on Data Placement Strategies for Cloud based Scientific Workflows, *141*(6), 30–33.

Slimani, S., Hamrouni, T., & Ben Charrada, F. (2021). Service-oriented replication strategies for improving quality-of-service in cloud computing: a survey. *Cluster Computing*, *24*(1), 361–392. https://doi.org/10.1007/s10586-020-03108-z

Souravlas, S., & Sifaleras, A. (2017a). Binary-tree based estimation of file requests for efficient data replication. *IEEE Transactions on Parallel and Distributed Systems*, *28*(7), 1839–1852. https://doi.org/10.1109/TPDS.2017.2650228

Souravlas, S., & Sifaleras, A. (2017b). On minimizing memory and computation overheads for binary-tree based data replication. *Proceedings - IEEE Symposium on Computers and Communications*, (1), 1296–1299. https://doi.org/10.1109/ISCC.2017.8024703

Spinnewyn, B., Botero, J. F., & Latre, S. (2018). Cost-effective replica management in fault-tolerant cloud environments. In *2017 13th International Conference on Network and Service Management, CNSM 2017* (Vol. 2018-Janua, pp. 1–9). https://doi.org/10.23919/CNSM.2017.8256019

Storey, V. C., & Song, I. Y. (2017). Big data technologies and Management: What conceptual modeling can do. *Data and Knowledge Engineering*, *108*(February), 50–67. https://doi.org/10.1016/j.datak.2017.01.001

Sun, D., Chang, G., Miao, C., & Wang, X. (2013). Analyzing, modeling and evaluating dynamic adaptive fault tolerance strategies in cloud computing environments. *Journal of Supercomputing*, *66*(1), 193–228. https://doi.org/10.1007/s11227-013-0898-7

Sun, D. W., Chang, G. R., Gao, S., Jin, L. Z., & Wang, X. W. (2012). Modeling a dynamic data replication strategy to increase system availability in cloud computing environments. *Journal of Computer Science and Technology*, *27*(2), 256–272. https://doi.org/10.1007/s11390-012-1221-4

Sun, S. Y., Yao, W. Bin, & Li, X. Y. (2018). DARS: A dynamic adaptive replica strategy under high load Cloud-P2P. *Future Generation Computer Systems*, *78*, 31–40. https://doi.org/10.1016/j.future.2017.07.046

Sureshpatil, D., Mane, R. V., & Ghorpade, V. R. (2018). Improving the Availability and Reducing Redundancy using Deduplication of Cloud Storage System. In *2017 International Conference on Computing, Communication, Control and Automation, ICCUBEA 2017* (pp. 1–5). https://doi.org/10.1109/ICCUBEA.2017.8463856

Tan, C. B., Hijazi, M. H. A., Lim, Y., & Gani, A. (2018). A survey on Proof of Retrievability for cloud data integrity and availability: Cloud storage state-of-the-art, issues, solutions and future trends. *Journal of Network and Computer Applications*, *110*(August 2017), 75–86. https://doi.org/10.1016/j.jnca.2018.03.017

Tchernykh, A., Schwiegelsohn, U., ghazali Talbi, E., & Babenko, M. (2016). Towards understanding uncertainty in cloud computing with risks of confidentiality, integrity, and availability. *Journal of Computational Science*. https://doi.org/10.1016/j.jocs.2016.11.011

Tchernykh, A., Schwiegelsohn, U., Talbi, E. ghazali, & Babenko, M. (2016). Towards understanding uncertainty in cloud computing with risks of confidentiality, integrity, and availability. *Journal of Computational Science*. https://doi.org/10.1016/j.jocs.2016.11.011

Tos, U., Mokadem, R., Hameurlain, A., & Ayav, T. (2021). Achieving query performance in the cloud via a cost-effective data replication strategy. *Soft Computing*, *25*(7), 5437–5454. https://doi.org/10.1007/s00500-020-05544-w

Tripathi, A., & Mishra, A. (2011). Cloud computing security considerations. *2011 IEEE International Conference on Signal Processing, Communications and Computing, ICSPCC 2011*. https://doi.org/10.1109/ICSPCC.2011.6061557

Vikhe, M., & Malhotra, J. (2015). A Review on Backup-up Practices using Deduplication, *4*(9), 338–344.

Wang, J., Wu, H., & Wang, R. (2017). A new reliability model in replication-based big data storage systems. *Journal of Parallel and Distributed Computing*, *108*, 14–27. https://doi.org/10.1016/j.jpdc.2017.02.001

Wang, L., Ma, Y., Yan, J., Chang, V., & Zomaya, A. Y. (2018). pipsCloud: High performance cloud computing for remote sensing big data management and processing. *Future Generation Computer Systems*, *78*, 353–368. https://doi.org/10.1016/j.future.2016.06.009

Wang, Y., & Wang, J. (2017). An Optimized Replica Distribution Method in Cloud Storage System. *Journal of Control Science and Engineering*, *2017*, 1–8. https://doi.org/10.1155/2017/2428982

Warhade, S., Dahiwale, P., & Raghuwanshi, M. M. (2016). A Dynamic Data Replication in Grid System. *Physics Procedia*, *78*(December 2015), 537–543. https://doi.org/10.1016/j.procs.2016.02.099

Watkins, K. (1993). *Discrete Event Simulation in C*. McGraw-Hill,International Series in Software Engineering.

Wehrle K., Güneş M., G. J. (2013). Introduction. In *Modeling and Tools for Network Simulation* (Vol. 21, pp. 1–2). Springer, Berlin, Heidelberg. Retrieved from https://doi.org/10.1007/978-3-642-12331-3_1

Weinman, J. (2017). The Evolving Cloud. *IEEE CLOUD COMPUTING*, 4–6.

Wu, X. (2016). Combination Replicas Placements Strategy for Data sets from Cost-effective View in the Cloud. *International Journal of Computational Intelligence Systems*, *10*(1), 521. https://doi.org/10.2991/ijcis.2017.10.1.36

Xia, Q., Liang, W., & Xu, Z. (2017). QoS-aware data replications and placements for query evaluation of big data analytics. *IEEE International Conference on Communications, 2017 IEEE*(2017), 1–7. https://doi.org/10.1109/ICC.2017.7997238

Xiao, Z., & Xiao, Y. (2013). Security and Privacy in Cloud Computing. *Communications Surveys Tutorials, IEEE*, *15*(2), 843–859. https://doi.org/10.1109/SURV.2012.060912.00182

Xie, F., Yan, J., & Shen, J. (2017). Towards Cost Reduction in Cloud-Based Workflow Management through Data Replication. *2017 Fifth International Conference on Advanced Cloud and Big Data (CBD)*, 94–99. https://doi.org/10.1109/CBD.2017.24

Xiong, L., Yang, L., Tao, Y., Xu, J., & Zhao, L. (2018a). Replication strategy for spatiotemporal data based on distributed caching system. *Sensors (Switzerland)*, *18*(1). https://doi.org/10.3390/s18010222

Xiong, L., Yang, L., Tao, Y., Xu, J., & Zhao, L. (2018b). Replication strategy for spatiotemporal data based on distributed caching system. *Sensors (Switzerland)*, *18*(1). https://doi.org/10.3390/s18010222

Xiong, R., Du, Y., Jin, J., Luo, J., Yang, W., Ciritoglu, H. E., … Li, H. (2018). A survey of dispatching rules for the dynamic unrelated machines environment. *Concurrency and Computation: Practice and Experience*, *XI*(Xi), 555–569. https://doi.org/10.1002/cpe.4830

Yan, Z., Zhang, L., Ding, W., & Zheng, Q. (2017). Heterogeneous Data Storage Management with Deduplication in Cloud Computing. *IEEE Transactions on Big Data*, (2), 1–1. https://doi.org/10.1109/tbdata.2017.2701352

Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth*, *10*(1), 13–53. https://doi.org/10.1080/17538947.2016.1239771

Yang, W., & Hu, Y. (2018). A replica management strategy based on MOEA/D. In *Proceedings of the 13th IEEE Conference on Industrial Electronics and Applications, ICIEA 2018* (pp. 2154–2159). IEEE. https://doi.org/10.1109/ICIEA.2018.8398067

Yi, M., Wei, J., & Song, L. (2017). Efficient integrity verification of replicated data in cloud computing system. *Computers and Security*, *65*, 202–212. https://doi.org/10.1016/j.cose.2016.11.003

Zeng, Z., & Veeravalli, B. (2014). Optimal metadata replications and request balancing strategy on cloud data centers. *Journal of Parallel and Distributed Computing*, *74*(10), 2934–2940. https://doi.org/10.1016/j.jpdc.2014.06.010

Zhang, Y., Nie, X., Jiang, J., Wang, W., Xu, K., Zhao, Y., … Yao, G. (2021). BDS+: An Inter-Datacenter Data Replication System With Dynamic Bandwidth Separation. *IEEE/ACM Transactions on Networking*, *29*(2), 918–934. https://doi.org/10.1109/TNET.2021.3054924

Zhao, Q., Xiong, C., Zhao, X., Yu, C., & Xiao, J. (2015). A data placement strategy for data-intensive scientific workflows in cloud. *Proceedings - 2015 IEEE/ACM 15th International Symposium on Cluster, Cloud, and Grid Computing, CCGrid 2015*, 928–934. https://doi.org/10.1109/CCGrid.2015.72