A Taxonomic Study: Data Placement Strategies in Cloud Replication Environments

Fazlina Mohd Ali[1], Marizuana Mat Daud[2], Fadilla Atyka Nor Rashid[3],
Nazhatul Hafizah Kamarudin[4], Syahanim Mohd Salleh[5], Nur Arzilawati Md Yunus[6]
Faculty of Information Science and Technology (FTSM), Universiti Kebangsaan Malaysia, Selangor, Malaysia[1, 3, 4, 5]
Institute of Visual Informatics, Universiti Kebangsaan Malaysia (UKM), Bangi 43600, Selangor, Malaysia[2]
Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM), Selangor, Malaysia[6]

***Abstract*—Since the past decades, the data replication trend has not subsided; it is progressing rapidly from multiple perspectives to enhance cloud replication performance. Researchers are eagerly focusing on improving the strategies in various perceptions; unfortunately, the vulnerability in every strategy is inevitable. A non-comprehensive replica strategy would have 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 consumption, and more, depending on the research areas. Many researchers are out of ideas to identify state-of-the-art issues. This exhaustive taxonomic study focused on analyzing the diversified contributions and limitations terrain of the cloud replication environment, focusing on data placement strategies. It seeks to delve deeply into its fundamental strategy, practical implementations, and the intricate challenges it poses. Concerning the imminent cloud-driven future, this structured review paper is a vital resource for researchers, policymakers, and industry professionals grappling with the complexities of this emerging paradigm. By illuminating the intricacies of data replication strategies, this study fosters a deeper appreciation for the transformative potential and the multifaceted challenges ahead of cloud data replications.**

Keywords*—Cloud environment; data replication; placement strategies; replication taxonomy; performance metrics

I. Introduction

An enormous amount of data is used extensively in the current era globally [1]–[3]. According to the International Data Corporation (IDC), global data will increase by 61% to 175 zettabytes by 2025, with most of that data being stored in cloud computing environments rather than data centers. This phenomenon is derived from the most interconnected Internet of Things (IoT) devices, leading to 100 billion terminals connected in 2025 [4], [5].

A. Cloud Computing

Cloud computing is well known as a data management platform that addresses the high volume of data demanded to be accessible by users anytime from anywhere [6], [7]. Cloud computing offers users a dynamic 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 [8]–[12].

Cloud computing offers users a “pay-as-you-go” basis since it enables multiple services, as shown in Fig. 1. The hardware resides as the fundamental facility in cloud computing architecture. The IaaS is the bottom layer in cloud services, which offers various large-scale infrastructure services with multi-specification of servers, CPU, memory, storage, and more. In the middle layer is PaaS, which enables numerous platforms like operating systems and software frameworks that can be tailored based on clients’ required environments. The top layer is SaaS, which is directly accessible to users with multiple applications such as web services, user interface systems like enterprise systems, and many more [13]–[15]. In the IaaS layer, the foremost benefit received by cloud users is the agility of the services. Cloud providers serve their clients by off-loading the hustle of managing the data center. Therefore, clients can freely focus on evolving their business [16].

Consequently, it improves the ability to meet user demands and reduce costs as the user can provision the resource amount accordingly [17]. Cost-saving is the most significant benefit a cloud tenant gains, as the pay-as-you-go paradigm reduces expenses on the overall data center maintenance cost. This foreseen cost can be a transition to operational expenses, vividly beneficial for business goals. Additionally, the SaaS layer delivers rapid deployment of client applications around the globe with only a few clicks [11]. PaaS enables the deployment of the necessary software applications, while SaaS, as the top layer, provides users with ready-to-use applications [13], [18]. These benefits have been the core reasons that drive users to choose the cloud as their data management platform.

B. Data Management

Data management is one of the prominent services enabled in cloud platforms. As a mass platform to serve high-volume data, the cloud is a multi-device technology that enables data management in a few deployment models: private cloud, public cloud, community cloud, and hybrid cloud. However, cloud tenants must choose an appropriate cloud model because every cloud model is distinguished depending on data criticality and resources [19]–[22]. Cloud providers need technical and business knowledge to propose the best model for organizations to ensure cloud tenants obtain efficient data management services. Comprehensive data management solutions are delivered to cloud tenants in the respective models, including data processing, security, storage, and recovery services.

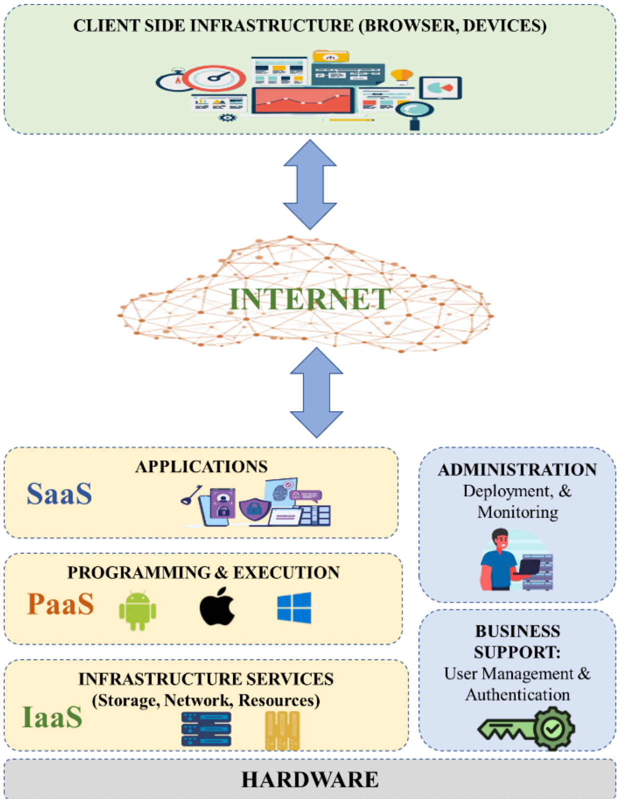

Fig. 1. Basic cloud computing architecture.

C. Data Recovery

Data Recovery must be adopted in any platform, including cloud environments [23]. The researcher in study [24] focused on data recovery in their study, whereby the researcher accentuated that the importance of data recovery is intolerable because the ambiguities of data absence are extremely anticipated in cloud platforms. Additionally, the researcher emphasizes that data recovery or failover costs are crucial before implementing cloud models. Therefore, the cloud as a data management platform has prepared data recovery mechanisms such as data backup, replication, and checkpointing [25]. The available approaches are playing different roles based on different circumstances.

D. Data Replication

Data replication is recognized as a promising cloud environment service that offers strategies to keep data in secure environments [25] safely. Data replication is defined as a heuristic multi-dimensional technique that saves one or more copies of data in multiple storage nodes in clouds [26]–[28]. According to study [29], data replication preserves the master data from catastrophic events (floods, earthquakes, etc.) and human errors, such as accidentally deleting information in the master files or deleting entire master files by users. The middleware manages replica copies in different environments known as disaster recovery centers. Therefore, copies of data are managed and kept safe in different nodes at different places. Thus, any unfortunate incidents that happen to the replicas would not affect the other replica copies. This would prevent data loss in any environment. Data replication is identified as a strategy that creates multiple copies in cloud storage in a big data environment, accelerating cloud system performance [29]–[31].

Replication strategies are commonly divided into two (2) mechanisms: static and dynamic data replications [32]–[34]. Static replication is a predetermined environment for dedicated cloud replication systems [14], [27]. The number of nodes, number of replicas, and many other architectures are fixed based on certain cloud system necessities before the replication strategy is implemented [35], [36]. The second mechanism is dynamic replication, or flexible replication strategies, where the algorithm can automatically create and delete replicas depending on the system users' access patterns [37]. However, the static replication mechanism is relatively simple and not preferable to be adapted in many cloud replication environments. The architecture is mainly foreordained, sometimes unsuitable for complex cloud replication systems [38], [39].

According to literature perceptions, a comprehensive cloud replication strategy always has a few significant phases; the first phase is usually the File Selection Strategy, which is to identify crucial data to be selected as replication candidates. The next phase is the Data Placement Technique, where the required number of replicas is identified, and the location to send the replicas is determined. Finally, the Data Center Selection Method is the stage to accomplish the replication process by identifying the best factors to select appropriate nodes to store the replicas in the cloud replication environment. Every stage complements different requirements. Researchers have the right to consolidate every phase in one research work as a complete replication process or develop every phase separately as fulfilments in respective research works. This research focuses on data placement strategies and factors in placing replicas in cloud environments.

A typical replication management architecture for a cloud environment is illustrated in Fig. 2.

Fig. 2 demonstrates a basic replication architecture in a cloud environment. Users in the architecture are also recognized as clients or tenants for cloud providers. The replication process is triggered when a user requests a data file from the cloud. A Global Replica Manager (GRM), typically the broker, manages and schedules replication tasks in the whole infrastructure. Conversely, the Local Replica Manager (LRM) manages local or inbound replication jobs for data centers. Usually, data centers are grouped in clusters, depending on the configuration of the cloud replication. The GRM and LRM must continually adhere to the rules and protocols in the algorithm specified in the replication system. The data file verification will occur when the manager receives user data requests. File information such as file names and locations are available as a comprehensive metadata table in the GRM and LRM. Then, the managers process the requested data according to the algorithm rules and identify candidate files for replication. The requested data file will be sent off to users, and new replica copies will be placed, conferring to the pre-determined placement technique in data centers.

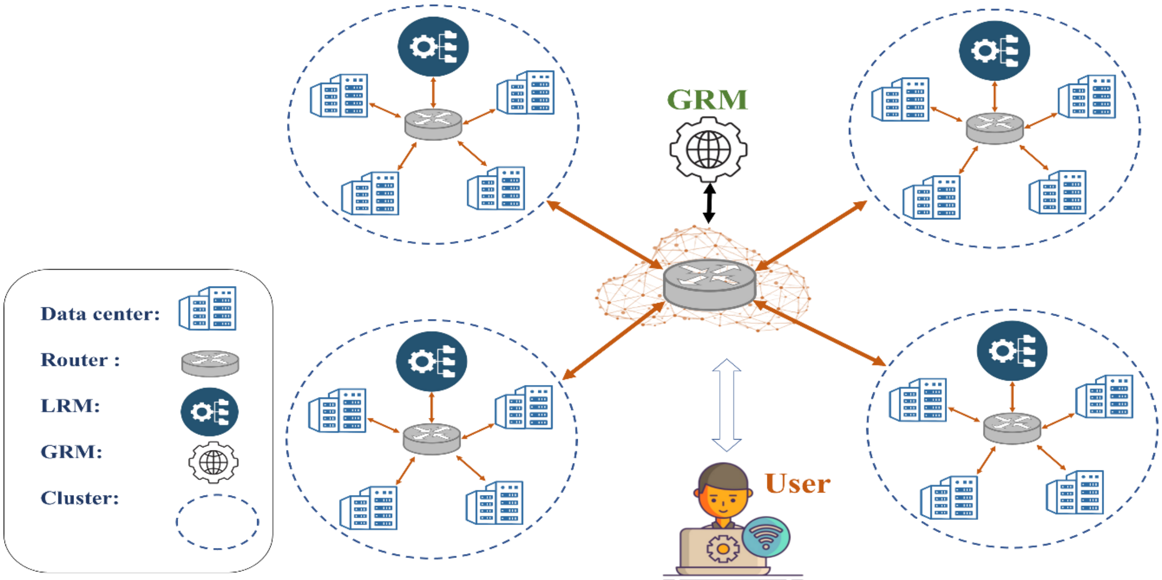

Fig. 2. Basic data replication architecture.

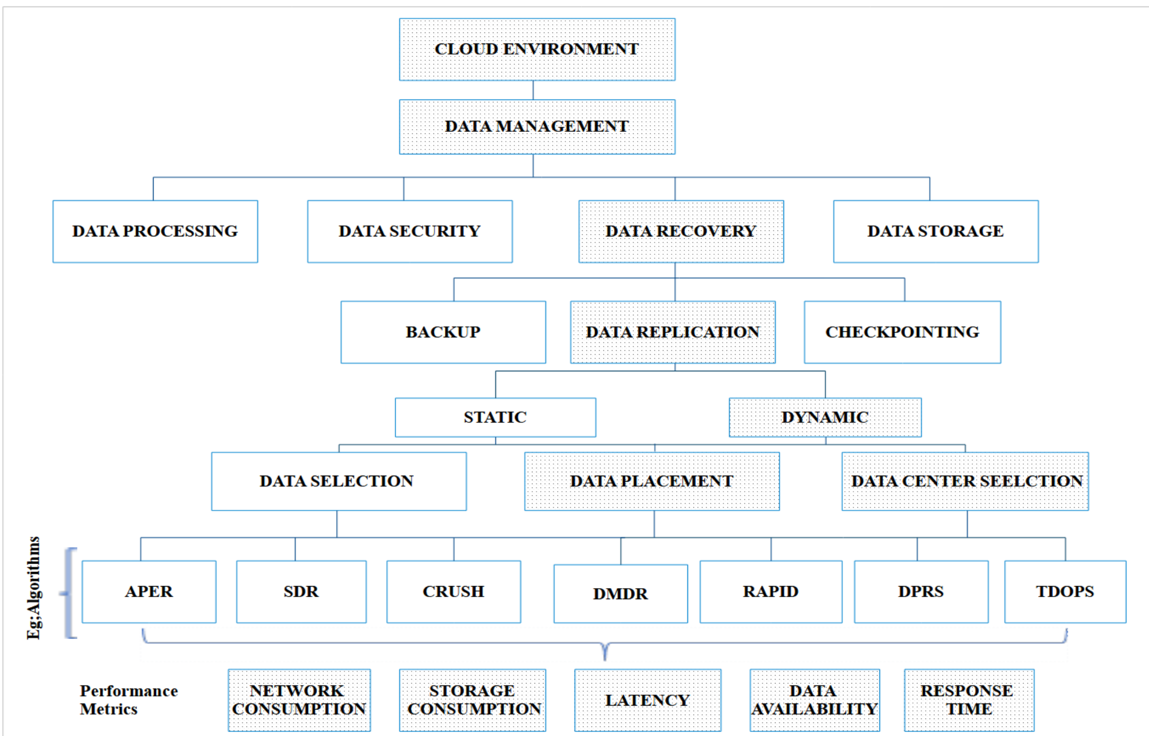

Fig. 3. Cloud replication taxonomy.

According to study [40] the taxonomy study can make the knowledge found in documents and texts clear and usable by other researchers also practitioners. Fig. 3 illustrates the data replication taxonomy in a cloud environment, which can be discovered in this study. This entire study is structured and organized as this taxonomy. The shaded boxes are the main focus of this study.

This taxonomy-based study thoroughly discusses similar studies related to this research area. The topic was explained and drilled down from top to bottom of the taxonomy, deriving the dedicated research area.

A taxonomy study on cloud computing is described, followed by the definition and overview of data management until the data replication body of knowledge.

A thorough discussion on existing research works in data replications and placements strategies with insights on contributions and achievements in numerous performance metrics enhancement.

Analysis tables and relevant analytics graphs are presented to share the essential details of related research works according to the research questions, highlighting research trends, contributions, and limitations.

II. Related Works

The advent of cloud computing has revolutionized how data is stored and managed. A critical aspect of cloud computing is data replication which involves creating and maintaining multiple copies of data across different locations to ensure high availability and reliability [41], [42].

Cloud computing empowers users with various resilient services that stand vibrantly as preferable technology for almost everyone to ensure data are efficiently managed and business continuity is guaranteed too [43]–[46]. 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. Fear of losing data during node failures is one of the core issues too [2], [11], [39], [47].

Hence, to mitigate the concerns that arise in a cloud platform, data replication is recognised as a promising cloud environment strategy [4], [7], [48]–[50]. Data replication is an empirical technique to accelerate system performance by generating identical data copies across multiple storage [51], [52]. Precisely, in a cloud environment, data replication is defined as creating several physical copies for every logical data item and locating the replica copies in different sites or storage nodes [36], [53], [54]. Depending on the cloud replication objectives, there are several ways to apply the data replication function.

There are two (2) prominent traditional techniques in cloud replication. First is static replication, where data copies are pre-determined and evenly distributed across nodes to ensure fault tolerance and load balancing. However, static replication may not adapt well to dynamic workloads and changing data access patterns. Dynamic replication dynamically adapts the replication scheme based on data access patterns and system conditions. These approaches aim to improve data availability and reduce access latency by dynamically creating or removing replicas as required. Dynamic replication techniques often utilize monitoring and feedback mechanisms to make informed decisions about replica placement [55], [56].

A taxonomy study is the structured names and definitions used to organize information and knowledge. Researcher in study [57] proposed a broad taxonomy of storage efficiency, with the performance metrics focused on cost optimization. The other research work by [58] introduces a taxonomy that organizes existing solutions for maintenance operations in cloud, edge, and IoT environments. This research work does not discuss any performance metrics yet; it merely reviews existing research work according to the taxonomy structure and presents the challenges within the research field. Similarly, [59] examines the migration field's characteristics and proposes a management-centric taxonomy in cloud computing. Researchers [60] and [61] embarked on cloud replication strategies and presented relevant taxonomy related to this research area. However, their studies proposed a very high-level taxonomy for this body of knowledge, which is not comprehensive enough to overview the entire structure of data replication processes in a cloud environment.

As enormous technologies are emerging extensively around the globe, many new approaches, such as artificial intelligence (AI), have been discovered as compatible techniques for replication strategies in cloud environments. Hence, researchers are continuously attempting novel replication strategies to place replica copies in cloud storage that provides multiple key services with resilient infrastructures for every cloud consumer [61], [62]. The cloud computing architectures, standards, and tools provide prospects for advancement in services, which offer various benefits to cloud clients [22], [63]–[65].

Thus, recent research works incorporated AI-based approaches in replication strategies to breed performance enhancements [66]. As such, similar studies like [67] and many state-of-the-art studies must not be overlooked and must be visible for future researchers' knowledge. In this context, there is still a lack of studies that present comprehensive surveys to produce a widespread taxonomy related to replication strategy in cloud environments. These limitations of the existing research motivated this study to produce a comprehensive taxonomy for data placement strategies in cloud replication environments. The taxonomy offers collections of replication strategies that allow cloud providers to scrutinize and implement them in real-cloud replication settings. Further, cloud providers would serve accelerated performance to users with better data availability, faster response time, low fault tolerance, reduced storage usage, and efficient network usage [28], [37], [68]–[70].

III. Method and Materials

This taxonomy study mainly explores the existing literature to investigate the coverage of multiple related topics, the research trends, and the critical review of relevant studies that have been published. The methodology implied in obtaining the source papers in this study mainly follows the guidelines suggested by [71]. According to study [40] the taxonomy study can make the knowledge found in documents and texts clear and usable by other researchers also practitioners. The guidelines for this study follow the essential steps of defining the research questions, searching for relevant papers, screening the papers, keywording the abstracts, extracting the data, and mapping. Each process step has an outcome, and the outcome of the complete process is the taxonomy mapping.

A. Research Questions (RQ)

1) Definition of research questions (Research Scope): The primary goal of a taxonomy study is to provide an overview of a research area and identify relevant research also results available within this field.

2) The Primary RQ of this study: 'What are the Data Placement Strategies in a Cloud Replication Environment?' This primary question was divided into four RQs. Table I lists the formulated RQs.

B. Data Sources

1) Search for primary studies (all papers): Primary research was found using keyword search terms on scientific databases or by personally looking through pertinent journal articles or conference proceedings.

2) *The primary data* Sources: Scopus online databases were primary data sources for potentially related studies. Other data sources were not considered to impede the overlapping of source results.

C. Search Terms

1) *Keywording of abstracts (classification scheme)*: Keywording is a way to reduce the time needed to develop the classification scheme and ensure that the method considers the existing studies.

2) *Search the related research effectively*: It is imperative to identify the pertinent search phrases. Kitchenham *et al.* [72] suggested population, intervention, comparison, and outcome (PICO) approach is fitting in this regard. Many review papers have broadly adopted these viewpoints. Here, the relevant PICO terms are listed:

- Population: Primary studies in Data Replication.
- Intervention: Placement strategies.
- Comparison: Strategies, Advantages, limitations.
- Performance metric, and future direction.
- Outcome: Placement Strategies, Advantages, and constraints in cloud replication environments.

D. Inclusion and Exclusion Criteria

1) *Screening of papers for inclusion and exclusion (relevant papers)*: Studies that were irrelevant to addressing the study issues were eliminated using inclusion and exclusion criteria.

2) *Inclusion and exclusion criteria*: applied to determine and discard relevant studies from the data sources to answer the RQs in this taxonomy study.

3) *Data extraction and taxonomy mapping of studies:* Once the classification scheme was in place, the relevant articles were discussed according to the structure.

E. Research Questions (RQ)

1) *Definition of research questions (research scope)*: The primary goal of a taxonomy study is to provide an overview of a research area and identify relevant research also results available within this field.

2) *The Primary RQ of this study:* 'What are the Data Placement Strategies in a Cloud Replication Environment?' This primary question was divided into four RQs. Table I lists the formulated RQs.

As shown in Fig. 4, a literature search was done using the keywords "(replication OR placement AND strategies AND in AND cloud), (replication AND strategies AND in AND cloud), (Data AND Replication OR placement AND strategies AND in AND cloud) identified 1,057 initial articles. A three-step screening process ensured high-quality and relevant studies. First, 399 articles were excluded for lacking peer-review or being in a language other than English. The second screening focused on article completeness, removing 399 articles lacking full text (available only as abstracts or presentations) or not directly relevant to cloud replication strategies. This left 258 articles for further evaluation. Finally, a rigorous selection process focusing on data placement strategies within cloud replication resulted in a final set of 25 anchor references for in-depth analysis. These principles were employed in all the studies retrieved during the different phases of the study selection procedure (see Table II)

TABLE I. RESEARCH QUESTIONS

RQ	Research Question	Motivation
RQ1.	What are the publication trends for research topics related to cloud data replication environments?	To investigate publication trends in this research field throughout recent years
RQ2.	What are data placement strategies and factors employed in existing research works?	To explore cloud replication strategies in state-of-the-art
RQ3.	What are the performance metrics that contribute to the enhancement of replica placement strategies in cloud replication environments?	To discover performance enhancements addressed in individual research
RQ4.	What are the common limitations across existing research on data placement strategies in cloud replication environments?	To gain the gaps in existing research that can guide future research explorations and improvements

TABLE II. INCLUSION AND EXCLUSION CRITERIA

Inclusion criteria	
IC1	Articles that are peer-reviewed
IC2	Articles providing research in Cloud Replication Strategies
IC3	Articles published from 2017 to 2024
Exclusion criteria	
EC1	Articles that do not meet the inclusion criteria
EC2	Articles without full text (only abstract or presentation)
EC3	Studies in languages other than English
EC4	Articles with unclear results or findings

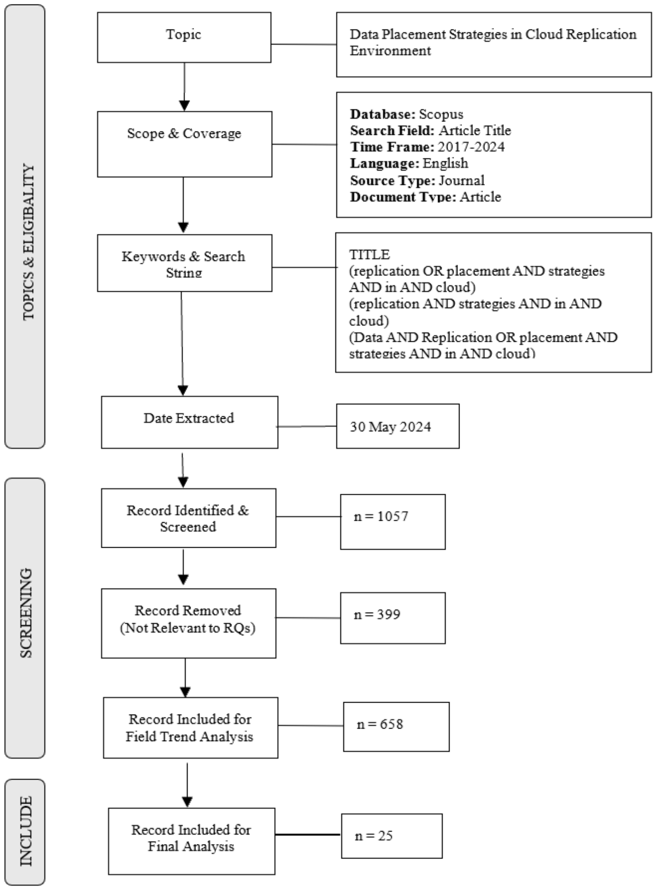

Fig. 4. Search strategy flow diagram.

IV. Findings and Discussions

A. *RQ1: What are the Publication Trends for Cloud Data Replication Environments Research Topics?*

Fig. 5 presents the publication trends in data placement strategies in cloud replication environments across six prominent journals and conferences from 2017 to May 2024. The data indicate significant growth in this research area, reflecting its increasing importance in cloud computing. The publication venues examined include Lecture Notes in Computer Science (LNCS), IEEE Access, Cluster Computing, Journal of Supercomputing, Future Generation Computer Systems (FGCS), and IEEE Transactions on Cloud Computing (TCC).

The total number of publications on data placement strategies in the Scopus online database from 2017 to May 2024 across all publication outlets is 658. The distribution of these publications among the six venues reveals distinct trends.

- Lecture Notes in Computer Science (LNCS) publications have increased from six papers in 2017 to 32 papers by May 2024. This represents a total of 175 publications, accounting for approximately 26.60% of the total publications.
- IEEE Access shows a significant rise from two papers in 2017 to 29 papers by May 2024, resulting in a total of 136 publications, which is about 20.67% of the total.
- Cluster Computing publications in *Cluster Computing* grew from 0 in 2017 to 26 by May 2024, with a total of 93 publications, making up around 14.14% of the total.
- Journal of Supercomputing starting from no publications in 2017 to 24 by May 2024, the *Journal of Supercomputing* accumulated a total of 110 publications, about 16.72% of the total.
- Future Generation Computer Systems (FGCS) publications increased from 3 in 2017 to 19 by May 2024, totaling 79 publications, approximately 12.01% of the total.
- IEEE Transactions on Cloud Computing (TCC) grew from no publications in 2017 to 19 by May 2024, amounting to a total of 65 publications, which is about 9.88% of the total.

The analysis of publication trends from 2017 to May 2024 illustrates a significant and growing interest in data placement strategies within cloud replication environments. This proves the evolving landscape in data placement strategies research, providing a foundation for future studies and developments in this pivotal area.

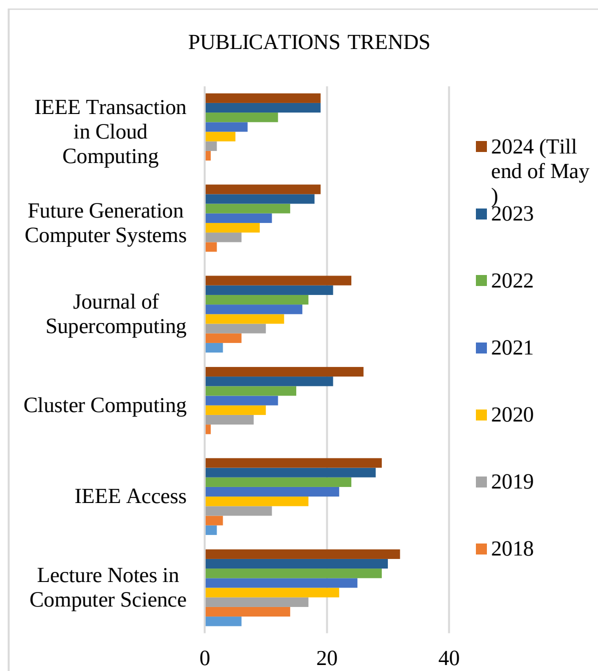

Fig. 5. Publication trends.

B. *RQ2: What Data Placement Strategies and Factors are Employed in Existing Research Works?*

1) Data placement strategies: Abundant studies cohesively developed numerous strategies to place replicas in the replication storage. At this point, scholars usually innovate a novel strategy to decide how to replicate the desired data, how many replicas are needed, and where to place the replicas [73]. the data placement method usually emphasizes the goal of accomplishing a cost-effective method with minimal storage consumption, high data availability, optimal replica copies, faster replication time, etc.

a) Low response time: A heuristic data replication and placements introduced by the authors [74] evaluated big data analytics queries in a distributed cloud. The researchers placed the source data of queries at multiple geo-distributed data centers. This technique ensured that the respective queries were locked with certain trial counts until they reached the specified threshold. The query will be passed to the next replica for a response. Another focus of this study is to place a sample of source data at appropriate data centers to meet users' rigorous query response time. The aim here is to minimise the evaluation cost while the response time is accelerated.

b) Low response time, low cost and low resource: Another group of researchers [75] developed a data replication strategy that mainly fulfils cloud tenants without neglecting cloud provider profits; the strategy was named Achieving Query Performance in the Cloud via a Cost-effective Data Replication (APER). The study proposed a cost-effective replica placement strategy that improves database queries with specific estimations to attain better response time. To achieve the desired response time, this study pre-determined that a particular replication time must be less than one particular service level objective (SLO) response time threshold. Later, replicas were placed using heuristic techniques that reduced both resource usage and monetary cost. The APER places new replicas by discarding previous copies from the cloud only if the threshold of access history is reached. As claimed by the researcher, the response time was successfully reduced in this study.

c) Low response time and low storage: The redundancy rate is another issue that can be rectified with comprehensive data replication. A group of researchers was dedicated and solved this problem by proposing a Time Series-based Deduplication and Optimal Data Placement Strategy (TDOPS) in their study [76]. Deduplication techniques were deployed, and data placement was determined based on capacity constraints, cloud data center load balance, and data transportation costs. The researchers in this study achieve improvements in space reduction, efficient retrieval, data transportation costs, and data transmission time.

d) High data availability: The Controlled Replication Under Scalable Hashing (CRUSH) algorithm has been improvised by the researcher [77] to resolve the bottleneck issues of the previous CRUSH. In the last CRUSH algorithm, replicas were inconsistently segregated in available storage nodes. The older version of CRUSH persistently sent generated replicas to the same active servers, consequently degrading the performance in the replication system due to a long queue to retrieve replicas, thus concurrently contributing to network congestion. The enhanced CRUSH algorithm [77] is recognised as a data placement technique capable of dynamically obtaining data at the next available replica place when the first requested replica is pending in response. In CRUSH, the proposed replication architecture to place data is the RING topology, where the direction to entertain user requests on data is more structured and bidirectional. As claimed in this research, the improved CRUSH method outstrips the previous studies in better data availability.

e) High data availability and low cost: Improving Clustering Based Critical Parent (CbCP) with a Replication (ICR) algorithm was proposed to address performance enhancement on replica placement scheduling [78]. The ICR consists of three sub-algorithms: scheduling algorithm, starting replication tasks, and task replication algorithm is used to identify any available resource until replica placement. The key aim of this study is to identify the possible and earliest time to start every replication task using the available resources without adding extra cost. Data reliability increased in this research, and execution costs were reduced significantly.

f) High data availability and low latency: A study by [79] offers an Artificial Bee Colony technique for data replication optimization in cloud environments. Another researcher suggests a multi-objective optimization data placement model based on numerous data replicas in a hybrid fog-cloud environment [80]. The articles suggest that AI can help minimize latency and improve data location.

g) Low cost and low latency: The researcher [81] proposed cost-effective, dynamic data placement, consisting of greedy and dynamic algorithms used to find the most reasonable cloud data placement solution. The greedy approach can find the optimum number of appropriate data centers to replicate the user's most accessed data. This study aims to deliver dynamic and optimised data placement with tolerable latency while incurring minimal service costs in social networks like Facebook. The proposed data placement strategy will determine the necessity of replica creation and choose the best data centers to place new replicas in the nearest data center for every user with minimal latency. The researcher aimed to reduce storage and latency, while the monetary problem was also addressed.

h) Low cost and low storage: Researcher in study [82] focused on financial cost reduction with good data consistency is the research objective attained by the researcher. The researcher proposed a Dynamic Replica Placement Strategy to satisfy the user experience while reducing the storage overheads, eventually contributing to cost reduction. The researcher implied a node renting concept to fulfill the capacity of the edge cloud to address the overload problem. Rented nodes are used whenever users in edge computing suffer low performance and release a rented node from the cloud during the users' stable performance. Therefore, this strategy evidenced that total financial costs were saved, and user experience was sustained during the replica placement phase.

i) Low cost, low storage usage, and efficient network: A dynamic data replication management technique was combined with a novel data placement strategy in the research work of [83]. The main goal was to reduce the network usage and costs associated with data transfers between data centers. Similarly, researchers [86] proposed Initialization Scheme-Based GA (ISGA) with a primary focus on reducing storage and network utilization costs. The study adopted an interval pricing technique to choose the best location for data in a multi-cloud environment.

j) Low storage: A Dynamic Redundant replica strategy based on the Security Level (SL-DRM) was proposed [84]. It flexibly adjusts the number of replicas and places the replica via constructing the data cache strategy using the Location Correlation of Cache (LC-Cache) to improve data read speed. The most crucial video footage was re-duplicated and saved in the storage. Therefore, as better security is required, the proposed technique dynamically changed the influence factors and instructed the algorithm to place new copies of video footage for a particular area. Thus, the data cache strategy focuses on the cameras' locations and time correlations of the video files by predicting the users' playback behaviors. This strategy achieved low storage consumption with a limited number of video copies.

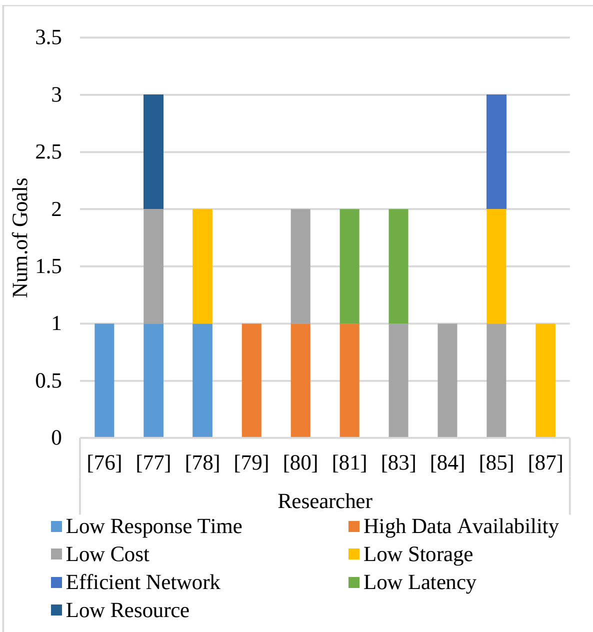

Fig. 6. Various goals achieved.

Fig. 6 is the bar chart that summarizes the key features and goals achieved by the various data placement strategies as discussed in this subsection. Each bar represents the number of goals achieved by a specific strategy. The strategies are listed along the x-axis, and the number of goals achieved is represented on the y-axis. The different colors indicate which specific goals are achieved by each strategy. This visual representation helps to quickly understand the focus and effectiveness of each strategy in achieving key performance and efficiency metrics.

2) Data placement strategies with data center selection factors: Data center selection is another significant method to place replicas in the cloud replication system process. This method is frequently integrated with the data placement process in almost every replication strategy. However, it is a considerable critical portion of accomplishing the replication process, whereby essential factors are determined to select the appropriate data center and storage to store replica copies. Therefore, researchers developed numerous methods to ensure a suitable location was identified. Usually, researchers determine a few factors to ensure the best data center has been determined to place the replica copies. The proposed factors or parameters directly affect various performance enhancements, in cloud replication environments [73]. Subsequent discussions will be categorized based on the number of factors employed in the respective research works.

a) Six factors: The Fuzzy Self-Defence Algorithm (FSDA) was proposed by study [85], which focused on a novel data center selection method. The FSDA determines the optimal number of replicas without degrading performance by implying a prey-predator model based on a fuzzy system to reproduce communication between prey and predator populations. The input parameters included in the fuzzy inference are system availability, service time, load, energy consumption, latency, and centrality. The researcher introduced several formulations to obtain merit values evaluated to determine the best nodes to place replica copies. The study claimed to improve hit ratio, energy consumption, and availability.

b) Five factors: A similar objective was achieved by [48] via producing a comprehensive algorithm named Cost Function based on an Analytic Hierarchy Process for data replication strategy (CF-AHP). CF-AHP is a multi-criteria optimization model that addresses cost-effective replica placement strategies that reduce energy consumption in data centers in cloud replication environments. The cost function is deployed to determine the best data center candidates to place newly generated replicas. The data center selection criterion consists of mean service time, access rate, latency, load variance, and storage usage. In the computation, weights are deployed to facilitate the user's task of determining system needs in respective parameters. The study achieved its goal of saving energy usage in its architecture.

In the same year, the author discovered an enhanced idea by proposing two (2) different Multiple-Criteria Decision Analysis (MCDA) approaches to determine the best data center to store replicas [86]. The first approach is to choose the best candidate site, and a cost function is computed using the weightage relationship concept in multi-criteria optimization known as AHP (Analytic Hierarchy Process). The second approach is ELECTRE-I, which consists of three (3) crucial stages; introduction of weight of criteria and calculation of concordance indices, calculation of the discordance indices, and over-ranking correlations. The criteria are further calculated using a weightage sum to obtain the values used to distinguish among available data centers. This study's data center selection criterion induces cost function calculation using AHP to acquire data center merits. The criterion consists of mean service time, access rate, latency, load variance, and storage usage, as adapted from another study by [87]. The respective criteria are computed using dedicated mathematical formulations. Results in this study evidenced that the cost function is sufficient to identify a candidate data center to place the replicas. The data center with the lowest cost value function is chosen to store the replica copies. Furthermore, when the system has insufficient space in storage nodes, replacement strategies are anticipated in this research work. This data center selection criteria method has attained efficient bandwidth usage and minimized data movement.

Another researcher developed a novel intelligent approach for dynamic data replication in a cloud environment [88]. The proposed algorithm in this research is bio-based. The first is the Multi-Objective Particle Swarm Optimization (MO-PSO), which selects a replica depending on the most requested. Second is the Ant Colony Optimization (MO-ACO) to retrieve the best replica placement decision. A data center selection criterion was used in this study using MO-ACO through comparing individual data centers based on the shortest distance, data centers with high access, storage capacity, and output, and data centers with a high number of hosts and VMs. All factors were compared, and the best data center with the highest data availability was selected to store replica copies. The study achieved betterments in replication cost by accelerating the response time and replication time and succeeded in enhancing network usage efficiency.

c) Four factors: In the same year, [67] proposed a CSO-based approach for Secure Data Replication (SDR) that uses fuzzy inferences to select the best data center to place replica copies. The data center selection criterion trained in the fuzzy inference is fewer than FSDA and centrality, energy, storage usage, and load. The best data center is selected via a fuzzy approach, and the data is chucked into a few segments to place in a few different storage nodes based on data center capacity. The parallel downloads can reduce download time and security because data files segregated in a few chunks in diverse data centers will not be meaningful if attackers compromise either one of the servers or data centers in cloud replication.

Another researcher focused on addressing optimization problems in a cloud replication environment [50]. The researcher enhanced the ant lion optimizer (ALO) algorithm and a fuzzy system by introducing a heuristic ALO (HH-ALO-Tabu). The proposed algorithm works dynamically in selecting the primary population based on chaotic maps (CMs), opposition-based learning (OBL), and random walk strategies depending on the differential evolution (DE) algorithm. The placement of replicas is based on four key parameters: service time, system availability, load variance, and providers' expenditures. With the key parameters, the selection of the best data center is determined, and replicas are placed accordingly. The study effectively guaranteed the cloud providers' economic revenue hands increased the users' satisfaction in upgrading the performance in cloud replication environments.

d) Three factors: Dynamic Popularity-aware Replication Strategy (DPRS) [89] constitutes a data center selection method to determine the best data center. The data center selection method is the final contribution to the research work, and it was implemented in every cluster with consideration of several significant factors. The factors are the number of file requests, storage availability, and data center distance. Each factor is computed to obtain average values in this phase, called data center merit. In this study, the weightage concept was adapted in the data center merit computation, whereby system administrator intervention was required to identify the necessary weights according to the system goals. Once the best data centers are determined, the candidate file is chunked to certain data block sizes and sent to a predetermined number of distributed storage nodes. DPRS attained efficient network usage with a parallel download concept and reduced replication frequency with the proposed data center selection method.

e) Two factors: Researchers proposed a replication placement and replacement technique with fuzzy-based deletion (HRS) for heterogeneous cloud data centers [90]. The study goal was to preserve storage space and enhance network efficiency. Thus, HRS ascertains the data center merit approach by considering a few significant parameters, such as the number of accesses and centrality with weightage for each parameter before the replica in storage nodes. The researcher selected a data center based on the temporal locality concept. Therefore, the data center with the highest access is considered in this study as closeness centrality. Subsequently, this researcher introduced the factors and computation to identify the lowest total cost provider. Additionally, HRS introduced fuzzy inference in the replacement strategy, which has insufficient space to store new replicas during storage. Therefore, the developed fuzzy algorithm will clarify existing files in other nodes and the last access history of a specific replica and predict future access to the replica. The researcher proved this by conducting beneficial experiments to address storage imbalance, insufficient response time, and high network usage.

Researchers [8] proposed Data Mining-based Data Replication (DMDR) in their research, which used data center selection criteria to select the best data center in the cloud replication environment. This study improvises storage utilization by introducing two (2) criteria; most central and the number of accesses. An accumulation using closeness formulation by [91] was adopted in DMDR. The weightage is used in the formulation, ensuring the system administrator has the authority to fulfill the system objective accordingly. Finally, if the data center does not have adequate storage space, the replacement strategy will be applied to delete unnecessary replicas based on predetermined factors. The researcher intended to minimise network usage during file retrieval by introducing this data center criterion.

Reducing the financial burden associated with excessive duplication presented by [54] with an approach to distribute Online Social Network (OSN) data across different Cloud Service Providers (CSPs). Their study presented an algorithm that uses metrics like access and interaction rates to determine which data objects are the most popular and in what order. The algorithm then calculates the minimum number of replica copies needed and places them in the best storage class. The storage of data is classified as Reduced Redundancy Storage (RRS), Standard Storage (SS), or Infrequent Access Storage (IAS), and distinct data are mapped to be placed in the most appropriate storage. Furthermore, when data becomes less often accessed over time, the system automatically switches to the Reduced Redundancy Storage (RRS) class by modifying the storage class dynamically based on access patterns. The study obtained low storage consumption with a minimum number of replica copies.

f) One factor: Research by [92] embarked on a dynamic replication approach that addresses massive data movement among data centers in the cloud. The author proposed an inter-data center data replication system with a Bandwidth Dynamic Separation algorithm called BDS+. The algorithm aims to speed up data transfer via adapting dynamic bandwidth separation, ensuring bandwidth is allocated for online traffic through estimating traffic demand and rescheduling bulk-data transfers for offline data services. It uses application-level multicast on the network with centralised architecture, which appoints the central controller to manage intermediate server data transmission. The researcher effectively improved bandwidth in this study.

The researcher proposed a cost-effective Hybrid PSO TS (HPSOTS) algorithm [93], which places restricted data copies in predetermined storage nodes to reduce energy consumption as the primary goal. The researcher used PSO's ability and TS's sturdy local search capability to obtain results on the appropriate data center to place replicas. Metaheuristic approaches were deployed to address the energy optimization issue via integrating Particle Swarm Optimization (PSO) and Tabu Search (TS) algorithms in their study. HPSOTS can determine the number of replicas before replication activities are triggered in a cloud environment. The integration of TS and PSO collaboratively exchanges inputs and eventually places the particles among six (6) fixed cloud data centers in encoded orders. The researchers succeeded in decreasing energy consumption and cost.

Conversely, [69] proposed the Replica Placement Based on Load Balance (RPBLB) technique to select the best data center to place replicas. The study aims to reduce remote users' access time. As replicas must always have fast retrieval rates, replicas must be placed in the closest nodes to users. Therefore, RPBLB most frequently duplicates access data in new storage nodes based on user demand for particular files. Research goals were achieved with response time reduced during user file download because data are kept in the closest data center.

An integrated algorithm between the Location-Aware Storage Technique (LAST) and the open-source Hadoop Distributed File System (HDFS) called LAST-HDFS was proposed by [94]. The study addressed the reliability of cloud providers and data integrity issues when users store their data in unknown storage locations. Unlike other research work, the LAST-HDFS did not use many parameters to determine the best data center to store data. On the other hand, the proposed algorithm allocated replicas to the data center by considering user-specified privacy policies. Users who store data in the same region place Every piece of data based on similar privacy preferences. The proposed algorithm proactively performs data placement balancing within the clusters and finally protects illegal data transfers through monitoring socket communications during migration or the replication process in the cloud environment. The study achieved high security by storing data according to privacy needs.

Table III summarizes the different strategies and number of factors considered for data placement in cloud replication environments. Fig. 7 depicts the factors employed in the data placement formulation to store the replica copies. No obvious trends suggest the number of factors considered has changed over time. Instead, to address some challenges, researchers have used these factors selectively, depending on their own research goals.

Holistically, every study shares similar research goals to improve performance in cloud replication environments, yet some ignored parallel drawbacks in their research. Existing studies contributed novel placement strategies for placing the replica copies in appropriate storage nodes. Table IV. includes a summary of existing research on data placement strategies, which comprises the contributions of every work.

TABLE III. SUMMARY OF DATA PLACEMENT STRATEGIES IN CLOUD REPLICATION

Strategy	Researcher Id	Number of Factors
FSDA	[85]	6
CF-AHP	[48]	5
MCDA	[86]	5
MO-ACO	[88]	5
HH-ALO-Tabu	[50]	4
SDR	[67]	4
DPRS	[89]	3
DMDR	[8]	2
HRS	[90]	2
OSN	[54]	2
LAST-HDFS	[94]	1
BDS+	[92]	1
HPSOTS	[93]	1
RPBLB	[69]	1

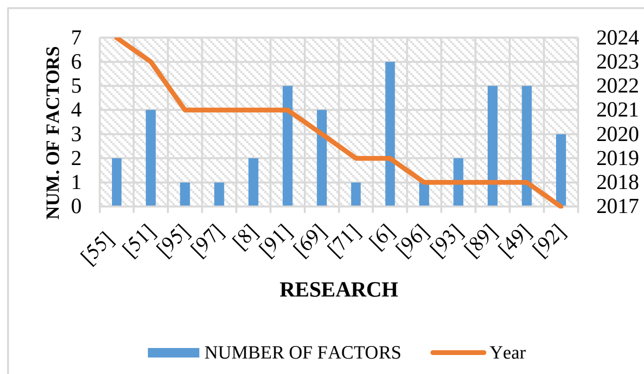

Fig. 7. Data placement factors.

C. *RQ3: What are the Performance Metrics that Contribute to the Enhancement of Replica Placement Strategies in Cloud Replication Environments?*

The existing research on data placement strategies in cloud replication environments has addressed various performance metrics advancement. As in Fig. 8, Efficient Network Usage (26%), Low Storage Usage (19%), and Low Response Time (18%) are the most significant advantages, collectively accounting for 63% of the total achievements. These metrics ensure efficient resource utilization, minimize storage costs, and provide timely data access in cloud replication environments.

TABLE IV. CONTRIBUTION DATA PLACEMENT STRATEGIES IN CLOUD REPLICATION

Ref.	Contribution Algorithm
[77]	Controlled Replication Under Scalable Hashing (CRUSH)
[74]	Heuristic Data Placement
[89]	Dynamic Popularity aware Replication Strategy (DPRS)
[84]	Security Level (SL-DRM)
[93]	Hybrid PSO TS (HPSOTS)
[90]	Hybrid data Replication Strategy (HRS)
[48]	Cost Function based on Analytic Hierarchy Process (CF-AHP)
[86]	Multiple-Criteria Decision Analysis (MCDA)
[69]	Replica Placement Based on Load Balance (RPBLB)
[8]	Data Mining-based Data Replication (DMDR)
[85]	Fuzzy Self-Defence Algorithm (FSDA)
[67]	Secure Data Replication (SDR)
[81]	Cost-Effective Dynamic Data Placement
[82]	Dynamic Replica Placement Strategy
[78]	Improving Clustering Based Critical Parent (CbCP)
[75]	Achieving Query Performance in the Cloud via Cost-effective Data Replication (APER)
[92]	Bandwidth Dynamic Separation (BDS+)
[94]	Location-Aware Storage Technique (LAST) and Hadoop Distributed File System (HDFS) called LAST-HDFS
[88]	Multi-Objective Particle Swarm Optimization (MO-PSO) and Ant Colony Optimization (MO-ACO)
[80]	Multi-objective optimization data placement model
[76]	Time Series-based Deduplication and Optimal Data Placement Strategy (TDOPS)
[50]	ALO (HH-ALO-Tabu)
[83]	Dynamic data replication and placement strategy
[95]	Initialization Scheme-Based GA (ISGA)
[54]	Distribute Online Social Network (OSN)

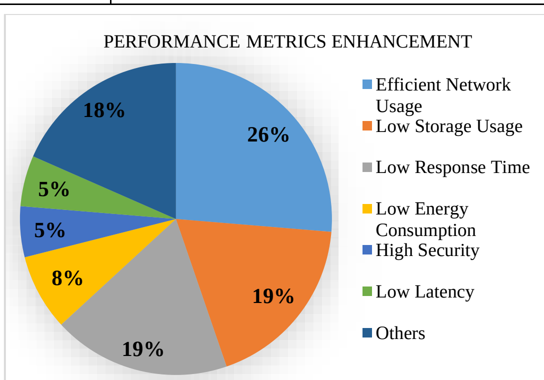

Fig. 8. Performance metrics.

Low Energy Consumption (8%), Low Latency, and High Security (5%) are also notable achievements. Other advantages of various performance metrics, such as low cost, high data availability, increased revenue, high user satisfaction, high load balance, and high hit ratio, are contributing to another 18% advancement. This stastical chart shows researchers' aims and successfully addresses various aspects of performance in cloud replication environments.

D. RQ4: What are the Common Limitations Across Existing Research on Data Placement Strategies in Cloud Replication Environments?

Here are the limitation in respective research works as discussed in RQ2:

- Researchers in [74] proposed a replica placement strategy that minimizes the evaluation cost. However, the disadvantage of this research is high bandwidth usage when queries are locked, and much hopping involved till the replicas are retrievable.
- The enhanced CRUSH algorithm [77] can dynamically place data and obtain data, which outstrips the previous study, yet data reliability issues are undeniable. The possibility of data loss is very high during data placement because when a dataset is sent to a particular server, any inevitable matters can happen at the server's level, such as server crashes or network breakdowns.
- SL-DRM achieved low storage consumption with a limited number of replica copies for videos [84]. Conversely, the major concern in this study was that the data availability was uncertain as crucial files were copied to unknown locations.

- The researcher [81] proposed cost-effective dynamic data placement using greedy algorithms. The solution was complex in formulation and caused extensive bandwidth usage while collecting numerous data, and prior decisions were made for replica creation and placement. Thus, computation overheads are a major drawback in this study.

- Researcher in study [82] proposed a Dynamic Replica Placement that focuses on financial cost reduction with good data consistency. As saving cost, storage space and consistency were the primary aims of this research, the data availability was neglected as it was not measured in this study.

- The ICR aims to identify the earliest available resource and time slots to proceed with replica replication until replica placement [78]. The trade-off in this study is high network usage due to the algorithm imposing an increased number of tasks to be executed whenever resources are available, and replica placement will occur endlessly. Furthermore, high energy consumption is another disadvantage in this study because resources are continuously used without idle time.

- Researchers in [75] developed the APER, a data replication strategy that allows a file to be replicated many times as long as it is profitable for both the tenant and cloud provider based on revenue and expenditures. The drawback is the high replication time affected by the overdue process of verifying every task to comply with the profitable criterion before the replication process and placement.

- In a hybrid Fog-Cloud environment, researcher Salah suggests a multi-objective optimization data placement model based on numerous data replicas [80]. Overall, the articles point to the possibility that while focusing on AI to minimize latency and improve data location the risks of data leakage concerning data privacy and the high requirement for specialist hardware and software are issues unattended in this study.

- The TDOPS with deduplication technique introduced in another study [76], overlooked the high replication time due to deduplication imposing a longer time for processing.

- Dynamic data replication management techniques notably focus on the costs related to storage and network use, ignoring the possibility of optimization through storage use [83].

- Researchers in study [95] adopted the ISGA, an interval pricing technique in multi-cloud environments, which has a primary disadvantage in the contributions of neglected cost-cutting strategies like resource optimization.

- The Dynamic Popularity-aware Replication Strategy (DPRS) [89] constitutes a data center selection method to determine the best location for replica placement. The researcher disregarded fault tolerance due to a heavy traffic load. Then the system will suffer from data loss and high response delay too.

- The HPSOTS algorithm was proposed by the researcher [93]. The data availability is relatively low due to the static number of replicas fixed in the strategy. Thus, any decisive data might not have sufficient copies, affecting poor availability and high wait time for file downloads.

- The researcher in study [90] proved the proposed HRS was beneficial to address storage imbalance, low response time, and high network usage. However, the system performance might suffer from execution overheads caused by extended verification time using the fuzzy technique.

- CF-AHP is a multi-criteria optimization model to save energy usage in their architecture [48]. The research disregarded the impact on the central database, which suffers a high update rate during replication.

- The Criteria Decision Analysis (MCDA) approach to determine the best data center to store replicas [86]. High energy usage and high replication time were neglected in this study.

- Researchers in study [71] propose the RPBLB strategy that requires additional storage to place replicas at the nearest data center. The algorithm demands extra time to verify the appropriate data center to store the replicas, thus affecting high process time.

- The additional computation time is identified in DMDR proposed by the researcher [8]. It has insufficient storage, and a high replication process while computing several factors during replica replacement activities that eventually delay the replication process.

- The researcher [85] and [67], introduced Fuzzy inferences in their placement strategies. Both researchers overlooked the fuzzy inference, causing a high process that derives a high response time. Concurrently, the proposed algorithm causes high storage consumption in [85].

- Research by [92] proposed an inter-data center data replication system that effectively improved bandwidth usage but overlooked high replication time. The algorithm requires more time to sort the traffic schedule than only triggering the replication process until the replica placement is complete.

- An integrated algorithm called LAST-HDFS was proposed by [94] achieved high security by storing data according to privacy classifications. However, the algorithm is cost-expensive due to the sophisticated security features. Another drawback in the research is high network usage caused by location monitoring and detection requiring data collection in real-time.

- Researchers [91] proposed a bio-based algorithm that overlooked the trade-off of computations overheads and high replication frequencies.

- Drawbacks in the heuristic ALO (HH-ALO-Tabu) strategy introduced by [50] is computational overheads.
- To decrease the financial liability associated with excessive duplication, [54] presented a strategy to distribute Online Social Network (OSN) data copies across a few Cloud Service Providers (CSPs). The study overlooked the impact of high costs and varying levels of data availability, which could present additional challenges when utilizing different services from various cloud service providers (CSPs).

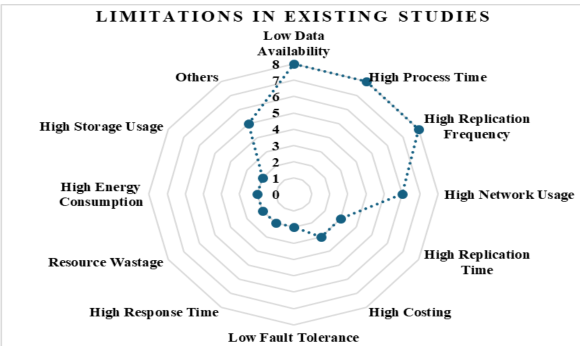

Fig. 9. Limitations in existing studies.

Fig. 9 and Table V present detailed analyses of common limitations in existing data placement strategies within cloud replication environments revealing several recurrent issues. The most frequently identified limitations include low data availability, high process time, and high replication frequency among 8 respective research studies (32%). These metrics indicate that many current strategies struggle to maintain consistent data access, handle data efficiently, and manage replication activities without excessive resource consumption. High network usage is down the road, with 6 research works contributing 24% of similar limitations. High-cost drawbacks (12%) are also significant concerns, highlighting inefficiencies in data transfer mechanisms and economic viability. These issues' occurrence emphasizes room for improvement and hinders performance degradation in replication activities.

TABLE V. RESEARCH WITH COMMON LIMITATIONS

Performance Metric	Authors
Low Data Availability	[77] [89] [93] [67] [82] [78] [54]
High Process Time	[90] [86] [85] [75] [92] [88] [80] [76] [50]
High Replication Frequency	[89] [84] [93] [8] [85] [92] [94] [88]
High Network Usage	[74] [85] [67] [81] [78] [94]
High Replication Time	[86] [75] [76]
High Costing	[69] [94] [80]
Low Fault Tolerance	[77] [89]
High Response Time	[48] [75]
Resource Wastage	[50] [83]
High Energy Consumption	[85] [78]
High Storage Usage	[89] [78]
Others	[89] [48] [75] [88] [80]

Numerous existing research works revealed that cloud computing has been the most prevalent platform for many other researchers' works [96]–[99]. Similarly, addressing these cloud replication limitations is critical for enhancing cloud computing environments' overall performance and reliability [100], [101]. Future research should prioritize the betterment of strategies that mitigate these common issues. By tackling these challenges, researchers can contribute to more robust and efficient data placement strategies, ultimately leading to more cost-effective and high-performing cloud services. Ensuring these improvements will be essential for meeting the increasing demands of cloud computing and providing reliable and accessible data management services.

V. CONCLUSION AND FUTURE DIRECTIONS

This paper presented a thorough taxonomic analysis of data placement strategies in cloud replication systems and has uncovered several important insights into this crucial area of cloud computing. This study has identified and categorized several techniques through this systematic review according to their primary goals, selection criteria for data placement, and performance metrics. The results suggest that placement strategies provide dynamic approaches that exhibit exceptional flexibility in response to shifting goal requirements. The study explicitly underscores how placement strategies have emerged as viable approaches to optimizing replica distribution in intricate cloud environments.

This study delivers a theoretical contribution that provides insights into existing research works by crafting a novel taxonomy for data placement strategies in cloud replication environments. The comprehensive analysis and discussion stipulate trends in replication performance enhancement. Further, gaps in existing studies are highlighted, forecasting new ideas for future researchers to develop a novel replication strategy to address the most prominent issues in cloud replication environments.

As for practical implications, this taxonomy study offers collections of replication strategies that allow cloud providers to scrutinize and adapt them in real-cloud replication settings. Hence, cloud providers can serve accelerated performance to users with better data availability, faster response time, low fault tolerance, reduced storage usage, and efficient network usage. This may offload pressure from cloud users' demands for quick data access, high data availability, fast replication process, low storage consumption, and affordable data maintenance.

Cloud technologies are evolving so quickly; thus, the taxonomy in this study may not fully encompass all new or specialized data placement techniques. Forthcoming tactics might differ greatly from those examined in this research. Future research can advance the field by addressing these gaps with more thorough literature discovery and provide practical solutions for efficient and dynamic data placement in cloud environments. Additionally, as the multi-cloud and hybrid cloud deployments grow more prevalent, data placement solutions that optimize across many cloud providers and on-premises infrastructure should be the focus of future studies. This is crafting strategies that can easily handle data placement in various distributed and heterogeneous situations. Besides, future researchers, policymakers, and professionals are suggested to explore more effective and dependable data placement, which could be ensured by utilizing artificial intelligence approaches to anticipate and adjust to changing circumstances. Real-world adaption may entail expanding and placing these ideas into practice by collaborating with businesses that use multi-cloud configurations.

ACKNOWLEDGMENT

This study is supported by the National University of Malaysia (UKM) (GGPM-2023-072). Sincerely grateful for the facilities and funding provided, which were essential for completing this research.

REFERENCES

[1] R. Kaur, I. Chana, and J. Bhattacharya, “Data deduplication techniques for efficient cloud storage management: a systematic review,” J. Supercomput., vol. 74, no. 5, pp. 2035–2085, 2018, doi: 10.1007/s11227-017-2210-8.

[2] S. Nannai John and T. T. Mirnalinee, “A novel dynamic data replication strategy to improve access efficiency of cloud storage,” Inf. Syst. E-bus. Manag., vol. 18, no. 3, pp. 405–426, 2020, doi: 10.1007/s10257-019-00422-x.

[3] M. Wang, Y. Qin, J. Liu, and W. Li, “Identifying personal physiological data risks to the Internet of Everything: the case of facial data breach risks,” Humanit. Soc. Sci. Commun., vol. 10, no. 1, pp. 1–15, 2023, doi: 10.1057/s41599-023-01673-3.

[4] Y. Shao, C. Li, Z. Fu, L. Jia, and Y. Luo, “Cost-effective replication management and scheduling in edge computing,” J. Netw. Comput. Appl., vol. 129, no. May 2018, pp. 46–61, 2019, doi: 10.1016/j.jnca.2019.01.001.

[5] K. Rajalakshmi, M. Sambath, L. Joseph, K. Ramesh, and R. Surendiran, “An Effective Approach for Improving Data Access Time using Intelligent Node Selection Model (INSM) in Cloud Computing Environment,” SSRG Int. J. Electr. Electron. Eng., vol. 10, no. 5, pp. 174–184, 2023, doi: 10.14445/23488379/IJEEE-V10I5P116.

[6] D. Mrozek, “A review of Cloud computing technologies for comprehensive microRNA analyses,” Comput. Biol. Chem., vol. 88, no. July, p. 107365, 2020, doi: 10.1016/j.compbiolchem.2020.107365.

[7] A. Shakarami, M. G. Ali, S. Mohammad, and M. Hamid, “Data Replication Schemes in Cloud Computing : A Survey,” Cluster Comput., vol. 7, 2021, doi: 10.1007/s10586-021-03283-7.

[8] N. Mansouri, M. M. Javidi, and B. Mohammad Hasani Zade, “Using data mining techniques to improve replica management in cloud environment,” Soft Comput., vol. 0123456789, 2019, doi: 10.1007/s00500-019-04357-w.

[9] C. Li, Y. Wang, H. Tang, and Y. Luo, “Dynamic multi-objective optimized replica placement and migration strategies for SaaS applications in edge cloud,” Futur. Gener. Comput. Syst., vol. 100, pp. 921–938, 2019, doi: 10.1016/j.future.2019.05.003.

[10] M. Vikhe and J. Malhotra, “A Review on Backup-up Practices using Deduplication,” vol. 4, no. 9, pp. 338–344, 2015.

[11] S. Logesswari, S. Jayanthi, D. Kalaiselvi, S. Muthusundari, and V. Aswin, “Materials Today : Proceedings A study on cloud computing challenges and its mitigations,” Mater. Today Proc., no. xxxx, 2020, doi: 10.1016/j.matpr.2020.10.655.

[12] U. Khalil, Mueen-Uddin, O. A. Malik, and S. Hussain, “A Blockchain Footprint for Authentication of IoT-Enabled Smart Devices in Smart Cities: State-of-the-Art Advancements, Challenges and Future Research Directions,” IEEE Access, vol. 10, no. June, pp. 76805–76823, 2022, doi: 10.1109/ACCESS.2022.3189998.

[13] A. Tripathi and A. Mishra, “Cloud computing security considerations,” 2011 IEEE Int. Conf. Signal Process. Commun. Comput. ICSPCC 2011, 2011, doi: 10.1109/ICSPCC.2011.6061557.

[14] N. Mansouri, R. Ghafari, and B. M. H. Zade, “Cloud computing simulators: A comprehensive review,” Simul. Model. Pract. Theory, vol. 104, no. July, 2020, doi: 10.1016/j.simpat.2020.102144.

[15] A. Arivuselvi, T. Amudha, P. Deepan Babu, and M. P. Scholar, “An Effective Ant Colony Optimization Methodology For Virtual Machine Placement In Cloud Data Centre,” turcomat.org, vol. 12, no. 10, pp. 3460–3467, 2021, Accessed: Aug. 07, 2023. [Online]. Available: https://www.turcomat.org/index.php/turkbilmat/article/view/5024

[16] F. A. M. Ibrahim and E. E. Hemayed, “Trusted Cloud Computing Architectures for infrastructure as a service: Survey and systematic literature review,” Comput. Secur., vol. 82, pp. 196–226, 2019, doi: 10.1016/j.cose.2018.12.014.

[17] T. Lynn, P. Rosati, A. Lejeune, and V. Emeakaroha, “A Preliminary Review of Enterprise Serverless Cloud Computing (Function-as-a-Service) Platforms,” Proc. Int. Conf. Cloud Comput. Technol. Sci. CloudCom, vol. 2017-Decem, pp. 162–169, 2017, doi: 10.1109/CloudCom.2017.15.

[18] V. C. Pezoulas, T. P. Exarchos, and D. I. Fotiadis, Cloud infrastructures for data sharing, no. iii. 2020. doi: 10.1016/b978-0-12-816507-2.00006-2.

[19] F. Lui et al., “NIST Cloud Computing Reference Architecture: Recommendations of NIST,” 2011.

[20] C. Ji, Y. Li, W. Qiu, U. Awada, and K. Li, “Big data processing in cloud computing environments,” Proc. 2012 Int. Symp. Pervasive Syst. Algorithms, Networks, I-SPAN 2012, pp. 17–23, 2012, doi: 10.1109/I-SPAN.2012.9.

[21] K. Grolinger, W. A. Higashino, A. Tiwari, and M. A. M. Capretz, “Data management in cloud environments: NoSQL and NewSQL data stores,” J. Cloud Comput., vol. 2, no. 1, 2013, doi: 10.1186/2192-113X-2-22.

[22] M. Yang, “Inverted Ant Colony Optimization Algorithm for Data Replication in Cloud Computing,” Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 7, pp. 1029–1038, 2023, doi: 10.14569/IJACSA.2023.01407111.

[23] V. Chang, “Towards a Big Data system disaster recovery in a Private Cloud,” Ad Hoc Networks, vol. 35, pp. 65–82, 2015, doi: 10.1016/j.adhoc.2015.07.012.

[24] M. M. Alshammari, A. A. Alwan, A. Nordin, and A. Z. Abualkishik, “Disaster Recovery with Minimum Replica Plan for Reliability Checking in Multi-Cloud,” Procedia Comput. Sci., vol. 130, pp. 247–254, 2018, doi: 10.1016/j.procs.2018.04.036.

[25] Y. Jun and Y. Lihong, “The Cloud Technology Double Live Data Center Information System Research and Design Based on Disaster Recovery Platform,” Procedia Eng., vol. 174, pp. 1356–1370, 2017, doi: 10.1016/j.proeng.2017.01.289.

[26] N. Chaturvedi, “Analysis of Replication and Replication Algorithms in Distributed System,” vol. 2, no. 5, pp. 261–266, 2012.

[27] S. George and E. B. Edwin, “A Review on Data Replication Strategy in Cloud Computing,” 2017 IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC 2017, pp. 1–4, 2018, doi: 10.1109/ICCIC.2017.8524269.

[28] M. Zheng, X. Du, Z. Lu, and Q. Duan, “A balanced and reliable data replica placement scheme based on reinforcement learning in edge–cloud environments,” Futur. Gener. Comput. Syst., vol. 155, no. November 2023, pp. 132–145, 2024, doi: 10.1016/j.future.2024.02.004.

[29] M. Shorfuzzaman and M. Masud, “Leveraging a multi-objective approach to data replication in cloud computing environment to support big data applications,” Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 3, pp. 418–429, 2019, doi: 10.14569/IJACSA.2019.0100354.

[30] S. S. Saleh, I. Alansari, M. K. Hamiaz, W. Ead, R. A. Tarabishi, and H. Khater, “iFogRep: An intelligent consistent approach for replication and placement of IoT based on fog computing,” Egypt. Informatics J., vol. 24, no. 2, pp. 327–339, 2023, doi: 10.1016/j.eij.2023.05.003.

[31] N. N. M.-R. O. B. D. R. in C. C. Vishnoi-I, V. Singh, S. Sharma, and P. Kumar, “Enhanced Naked Mole-Rat Optimization Based Data Replication in Cloud Computing,” Int. J. Intell. Syst. Appl. Eng., vol. 12, no. 3s, pp. 355 – 362, 2024, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85179340068&partnerID=40&md5=924d44ff456f26dcb0fef55fdfbb7d55

[32] R. Mokadem and A. Hameurlain, “A data replication strategy with tenant performance and provider economic profit guarantees in Cloud data centers,” J. Syst. Softw., vol. 159, p. 110447, 2020, doi: 10.1016/j.jss.2019.110447.

[33] M. Seguela, R. Mokadem, and J.-M. Pierson, "Comparing energy-aware vs. cost-aware data replication strategy," Int. Green Sustain. Comput. Conf., pp. 1–8, 2020, doi: 10.1109/igsc48788.2019.8957206.

[34] B. Arasteh, S. S. Sefati, S. Halunga, O. Fratu, and T. Allahviranloo, "A Hybrid Heuristic Algorithm Using Artificial Agents for Data Replication Problem in Distributed Systems," Symmetry (Basel)., vol. 15, no. 2, 2023, doi: 10.3390/sym15020487.

[35] N. K. Gill and S. Singh, "A dynamic, cost-aware, optimized data replication strategy for heterogeneous cloud data centers," Futur. Gener. Comput. Syst., vol. 65, pp. 10–32, 2016, doi: 10.1016/j.future.2016.05.016.

[36] B. Alami Milani and N. Jafari Navimipour, "A comprehensive review of the data replication techniques in the cloud environments: Major trends and future directions," J. Netw. Comput. Appl., vol. 64, pp. 229–238, 2016, doi: 10.1016/j.jnca.2016.02.005.

[37] B. Alami Milani and N. Jafari Navimipour, "A Systematic Literature Review of the Data Replication Techniques in the Cloud Environments," Big Data Res., vol. 10, no. C, pp. 1–7, 2017, doi: 10.1016/j.bdr.2017.06.003.

[38] C. Li, Y. Wang, H. Tang, Y. Zhang, Y. Xin, and Y. Luo, "Flexible replica placement for enhancing the availability in edge computing environment," Comput. Commun., 2019, doi: 10.1016/j.comcom.2019.07.013.

[39] K. Sabaghian, K. Khamforoosh, and A. Ghaderzadeh, "Data Replication and Placement Strategies in Distributed Systems: A State of the Art Survey," Wireless Personal Communications, vol. 129, no. 4. Springer, pp. 2419–2453, Apr. 01, 2023. doi: 10.1007/s11277-023-10240-7.

[40] S. Yazdani, A. Shirvani, and P. Heidarpoor, "A Model for the Taxonomy of Research Studies: A Practical Guide to Knowledge Production and Knowledge Management," Archives of Pediatric Infectious Diseases, vol. 9, no. 4. 2021. doi: 10.5812/pedinfect.112456.

[41] P. Elango, K. Kuppusamy, and N. Prabhu, "Data Replication Using Data Mining Techniques," Asian J. Comput. Sci. Technol., vol. 8, no. S1, pp. 107–109, 2021, doi: 10.51983/ajcst-2019.8.s1.1939.

[42] I. Kuraj, A. Solar-Lezama, and N. Polikarpova, "POSTER: Optimizing Consistency for Partially Replicated Data Stores," Proc. ACM SIGPLAN Symp. Princ. Pract. Parallel Program. PPOPP, pp. 457–458, 2022, doi: 10.1145/3503221.3508438.

[43] C. B. Tan, M. H. A. Hijazi, Y. Lim, and A. Gani, "A survey on Proof of Retrievability for cloud data integrity and availability: Cloud storage state-of-the-art, issues, solutions and future trends," J. Netw. Comput. Appl., vol. 110, no. August 2017, pp. 75–86, 2018, doi: 10.1016/j.jnca.2018.03.017.

[44] N. K. Nivetha and D. Vijayakumar, "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, IEEE, 2016, pp. 1–6. doi: 10.1109/ICCTIDE.2016.7725322.

[45] Q. Liu, G. Wang, and J. Wu, "Consistency as a service: Auditing cloud consistency," IEEE Trans. Netw. Serv. Manag., vol. 11, no. 1, pp. 25–35, 2014, doi: 10.1109/TNSM.2013.122613.130411.

[46] R. D and G. A, "Optimization assisted frequent pattern mining for data replication in cloud: Combining sealion and grey wolf algorithm," Adv. Eng. Softw., vol. 176, no. January, p. 103401, 2023, doi: 10.1016/j.advengsoft.2022.103401.

[47] J. Wang, H. Wu, and R. Wang, "A new reliability model in replication-based big data storage systems," J. Parallel Distrib. Comput., vol. 108, pp. 14–27, 2017, doi: 10.1016/j.jpdc.2017.02.001.

[48] M. R. Djebbara and H. Belbachir, "Cost Function Based On Analytic Hierarchy Process for Data Replication Strategy in Cloud Environment," J. Theor. Appl. Inf. Technol., vol. 96, pp. 2638–2648, 2018.

[49] F. Xie, J. Yan, and J. Shen, "Towards Cost Reduction in Cloud-Based Workflow Management through Data Replication," 2017 Fifth Int. Conf. Adv. Cloud Big Data, pp. 94–99, 2017, doi: 10.1109/CBD.2017.24.

[50] B. Mohammad Hasani Zade, N. Mansouri, and M. M. Javidi, A new hyper-heuristic based on ant lion optimizer and Tabu search algorithm for replica management in cloud environment, vol. 56, no. 9. Springer Netherlands, 2023. doi: 10.1007/s10462-022-10309-y.

[51] B. Spinnewyn, J. F. Botero, and S. Latre, "Cost-effective replica management in fault-tolerant cloud environments," in 2017 13th International Conference on Network and Service Management, CNSM 2017, 2018, pp. 1–9. doi: 10.23919/CNSM.2017.8256019.

[52] N. Mansouri, M. M. Javidi, and B. M. H. Zade, "Hierarchical data replication strategy to improve performance in cloud computing," Front. Comput. Sci., vol. 15, no. 2, 2021, doi: 10.1007/s11704-019-9099-8.

[53] A. Khelifa, T. Hamrouni, R. Mokadem, and F. Ben Charrada, "SLA-aware task scheduling and data replication for enhancing provider profit in clouds," Procedia Comput. Sci., vol. 176, pp. 3143–3152, 2020, doi: 10.1016/j.procs.2020.09.174.

[54] A. Y. Aldailamy, A. Muhammed, R. Latip, N. A. W. A. Hamid, and W. Ismail, "Online dynamic replication and placement algorithms for cost optimization of online social networks in two-tier multi-cloud," J. Netw. Comput. Appl., vol. 224, p. 103827, 2024, doi: 10.1016/j.jnca.2024.103827.

[55] M. A. Fazlina, R. Latip, M. A. Alrshah, and S. Member, "Vigorous Replication Strategy with Balanced Quorum for Optimizing The Storage and Response Time in Cloud Environments," pp. 1–13, 2021.

[56] P. Zhao, X. Sun, J. Shang, J. Lin, M. Dong, and B. Li, "A Dynamic Convergent Replica Selection Strategy Based on Cloud Storage," Proc. - 2019 Int. Conf. Artif. Intell. Adv. Manuf. AIAM 2019, pp. 473–478, 2019, doi: 10.1109/AIAM48774.2019.00100.

[57] M. Liu, L. Pan, and S. Liu, "Cost Optimization for Cloud Storage from User Perspectives: Recent Advances, Taxonomy, and Survey," ACM Comput. Surv., vol. 55, no. 13s, Jul. 2023, doi: 10.1145/3582883.

[58] P. Souza, T. Ferreto, and R. Calheiros, "Maintenance Operations on Cloud, Edge, and IoT Environments: Taxonomy, Survey, and Research Challenges," ACM Comput. Surv., vol. 56, no. 10, Jun. 2024, doi: 10.1145/3659097.

[59] T. He and R. Buyya, "A Taxonomy of Live Migration Management in Cloud Computing," ACM Comput. Surv., vol. 56, no. 3, Oct. 2023, doi: 10.1145/3615353.

[60] Q. Waseem, W. I. S. Wan Din, S. S. Alshamrani, A. Alharbi, and A. Nazir, "Quantitative analysis and performance evaluation of target-oriented replication strategies in cloud computing," Electron., vol. 10, no. 6, pp. 1–49, 2021, doi: 10.3390/electronics10060672.

[61] S. Slimani, T. Hamrouni, and F. Ben Charrada, "Service-oriented replication strategies for improving quality-of-service in cloud computing: a survey," Cluster Comput., vol. 24, no. 1, pp. 361–392, 2021, doi: 10.1007/s10586-020-03108-z.

[62] J. Ma, G. Wang, and X. Liu, "DedupeSwift: Object-oriented storage system based on data deduplication," in IEEE Trustcom/BigDataSE/ISPA, IEEE, 2016, pp. 1069–1076. doi: 10.1109/TrustCom.2016.0177.

[63] L. Rajabion, A. A. Shaltooki, M. Taghikhah, A. Ghasemi, and A. Badfar, "Healthcare big data processing mechanisms: The role of cloud computing," Int. J. Inf. Manage., vol. 49, no. June 2017, pp. 271–289, 2019, doi: 10.1016/j.ijinfomgt.2019.05.017.

[64] D. L. J. C.Venish Raja, "A Cost Effective Scalable Scheme for Dynamic Data Service in Heterogeneous Cloud Environment," Int. J. Adv. Sci. Technol., vol. Vol. 28, N, no. January, 2020.

[65] B. Prasad Babu, P. Peddi, and S. Pappala, "Optimization for Dynamic Replication in Cloud Center," Int. J. Adv. Res. Arts, Sci. Eng. Manag., vol. 10, no. 5, pp. 2641–2646, 2023.

[66] M. A. Ahmed, M. H. Khafagy, M. E. Shaheen, and M. R. Kaseb, "Dynamic Replication Policy on HDFS Based on Machine Learning Clustering," IEEE Access, vol. 11, no. January, pp. 18551–18559, 2023, doi: 10.1109/ACCESS.2023.3247190.

[67] N. Mansouri, M. M. Javidi, and B. A. Mohammad Hasani Zade, "A CSO-based approach for secure data replication in cloud," J. Supercomput., vol. 77, no. 6, pp. 5882–5933, 2020, doi: 10.1007/s11227-020-03497-3.

[68] S. U. R. Malik et al., "Performance analysis of data intensive cloud systems based on data management and replication: a survey," Distrib. Parallel Databases, vol. 34, no. 2, pp. 179–215, 2016, doi: 10.1007/s10619-015-7173-2.

[69] X. Fu, J. Li, W. Liu, S. Deng, and J. Wang, "Data Replica Placement Policy Based on Load Balance in Cloud Storage System," 2019 IEEE 3rd

Inf. Technol. Networking, Electron. Autom. Control Conf., no. Itnec, pp. 682–685, 2019, doi: 10.1109/ITNEC.2019.8728995.

[70] K. Liu, J. Peng, J. Wang, W. Liu, Z. Huang, and J. Pan, "Scalable and adaptive data replica placement for geo-distributed cloud storages," IEEE Trans. Parallel Distrib. Syst., vol. 31, no. 7, pp. 1575–1587, 2020, doi: 10.1109/TPDS.2020.2968321.

[71] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," 12th Int. Conf. Eval. Assess. Softw. Eng. EASE 2008, no. June, 2008, doi: 10.14236/ewic/ease2008.8.

[72] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering - A systematic literature review," Information and Software Technology, vol. 51, no. 1. Elsevier B.V., pp. 7–15, 2009. doi: 10.1016/j.infsof.2008.09.009.

[73] S. Mazumdar, D. Seybold, K. Kritikos, and Y. Verginadis, A survey on data storage and placement methodologies for Cloud-Big Data ecosystem, vol. 6, no. 1. Springer International Publishing, 2019. doi: 10.1186/s40537-019-0178-3.

[74] Q. Xia, W. Liang, and Z. Xu, "QoS-aware data replications and placements for query evaluation of big data analytics," IEEE Int. Conf. Commun., vol. 2017 IEEE, no. 2017, pp. 1–7, 2017, doi: 10.1109/ICC.2017.7997238.

[75] U. Tos, R. Mokadem, A. Hameurlain, and T. Ayav, "Achieving query performance in the cloud via a cost-effective data replication strategy," Soft Comput., vol. 25, no. 7, pp. 5437–5454, 2021, doi: 10.1007/s00500-020-05544-w.

[76] S. U. Muthunagai and R. Anitha, "TDOPS: Time series based deduplication and optimal data placement strategy for IIoT in cloud environment," J. Intell. Fuzzy Syst., vol. 43, no. 1, pp. 1583–1597, 2022, doi: 10.3233/JIFS-212568.

[77] P. Carns, K. Harms, J. Jenkins, M. Mubarak, R. Ross, and C. Carothers, "Impact of data placement on resilience in large-scale object storage systems," IEEE Symp. Mass Storage Syst. Technol., 2017, doi: 10.1109/MSST.2016.7897091.

[78] S. S. Mousavi Nik, M. Naghibzadeh, and Y. Sedaghat, "Task replication to improve the reliability of running workflows on the cloud," Cluster Comput., vol. 24, no. 1, pp. 343–359, 2021, doi: 10.1007/s10586-020-03109-y.

[79] R. Salem, M. A. Salam, H. Abdelkader, and A. Awad Mohamed, "An Artificial Bee Colony Algorithm for Data Replication Optimization in Cloud Environments," IEEE Access, vol. 8, pp. 51841–51852, 2020, doi: 10.1109/ACCESS.2019.2957436.

[80] N. Ben Salah and N. Bellamine Ben Saoud, An IoT-oriented Multiple Data Replicas Placement Strategy in Hybrid Fog-Cloud Environment, vol. 1, no. 1. Association for Computing Machinery, 2021. doi: 10.1145/3437959.3459251.

[81] H. Khalajzadeh, D. Yuan, B. B. Zhou, J. Grundy, and Y. Yang, "Cost effective dynamic data placement for efficient access of social networks," J. Parallel Distrib. Comput., vol. 141, pp. 82–98, 2020, doi: 10.1016/j.jpdc.2020.03.013.

[82] C. Li, J. Bai, Y. Chen, and Y. Luo, "Resource and replica management strategy for optimizing financial cost and user experience in edge cloud computing system," Inf. Sci. (Ny)., vol. 516, pp. 33–55, 2020, doi: 10.1016/j.ins.2019.12.049.

[83] L. Bouhouch, M. Zbakh, and C. Tadonki, "Dynamic data replication and placement strategy in geographically distributed data centers," in Concurrency and Computation: Practice and Experience, 2023. doi: 10.1002/cpe.6858.

[84] R. Li, J. Zhang, and W. Shen, "Replicas Strategy and Cache Optimization of Video Surveillance Systems Based on Cloud Storage," Futur. Internet, vol. 10, no. 4, p. 34, 2018, doi: 10.3390/fi10040034.

[85] N. Mansouri, B. Mohammad Hasani Zade, and M. M. Javidi, "A multi-objective optimized replication using fuzzy based self-defense algorithm for cloud computing," J. Netw. Comput. Appl., vol. 171, no. October 2019, p. 102811, 2020, doi: 10.1016/j.jnca.2020.102811.

[86] M. R. Djebbara and H. Belbachir, "Cloud data replication strategy using multiple-criteria decision analysis methods," Multiagent Grid Syst., vol. 14, no. 2, pp. 203–218, 2018, doi: 10.3233/mgs-180288.

[87] M. Radi, "Replica Placement Strategy Based on Analytic Hierarchy Process in Heterogeneous Cloud Data Storage," Int. Arab Conf. Inf. Technol. (ACIT ' 2016), no. December 2016, 2017.

[88] A. Awad, R. Salem, H. Abdelkader, and M. A. Salam, "A Novel Intelligent Approach for Dynamic Data Replication in Cloud Environment," IEEE Access, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3064917.

[89] N. Mansouri, M. K. Rafsanjani, and M. M. Javidi, "DPRS: A dynamic popularity aware replication strategy with parallel download scheme in cloud environments," Simul. Model. Pract. Theory, vol. 77, pp. 177–196, 2017, doi: 10.1016/j.simpat.2017.06.001.

[90] N. Mansouri and M. M. Javidi, "A hybrid data replication strategy with fuzzy-based deletion for heterogeneous cloud data centers," J. Supercomput., vol. 74, no. 10, pp. 5349–5372, 2018, doi: 10.1007/s11227-018-2427-1.

[91] M. E. J. Newman, "Networks of networks - An introduction," (2010, Oxford Univ. Press. Artif. life., vol. 80, pp. 1–6, 2009, doi: 10.1016/j.chaos.2015.03.016.

[92] Y. Zhang et al., "BDS+: An Inter-Datacenter Data Replication System With Dynamic Bandwidth Separation," IEEE/ACM Trans. Netw., vol. 29, no. 2, pp. 918–934, 2021, doi: 10.1109/TNET.2021.3054924.

[93] Y. Ebadi and N. Jafari Navimipour, "An energy-aware method for data replication in the cloud environments using a Tabu search and particle swarm optimization algorithm," Concurr. Comput. Pract. Exp., no. July 2017, p. e4757, 2018, doi: 10.1002/cpe.4757.

[94] A. Bowers, C. Liao, D. Steiert, D. Lin, A. Squicciarini, and A. Hurson, "Detecting Suspicious File Migration or Replication in the Cloud," IEEE Trans. Dependable Secur. Comput., vol. Vol. 18, N, no. 1, pp. 296–309, 2021.

[95] P. Wang, Z. Chen, M. C. Zhou, Z. Zhang, A. Abusorrah, and A. C. Ammari, "Cost-Effective and Latency-Minimized Data Placement Strategy for Spatial Crowdsourcing in Multi-Cloud Environment," IEEE Trans. Cloud Comput., vol. 11, no. 1, pp. 868–878, 2023, doi: 10.1109/TCC.2021.3119862.

[96] H. M. F. Noman et al., "Machine Learning Empowered Emerging Wireless Networks in 6G: Recent Advancements, Challenges and Future Trends," IEEE Access, vol. 11, no. July, pp. 83017–83051, 2023, doi: 10.1109/ACCESS.2023.3302250.

[97] N. K. Naseri, E. Sundararajan, and M. Ayob, "Smart Root Search (SRS) in Solving Service Time–Cost Optimization in Cloud Computing Service Composition (STCOCCSC) Problems," Symmetry (Basel)., vol. 15, no. 2, Feb. 2023, doi: 10.3390/sym15020272.

[98] F. Mohd Ali, N. A. Md Yunus, N. N. Mohamed, M. Mat Daud, and E. A. Sundararajan, "A Systematic Mapping: Exploring Internet of Everything Technologies and Innovations," Symmetry, vol. 15, no. 11. p. 1964, 2023. doi: 10.3390/sym15111964.

[99] L. S. Xue, N. A. A. Majid, and E. A. Sundararajan, "A principal component analysis and clustering based load balancing strategy for cloud computing," TEM J., vol. 9, no. 1, pp. 93–100, 2020, doi: 10.18421/TEM91-14.

[100] S. S. Alnuaimi, E. A. Sundararajan, and A. H. A. Rahman, "Data Distribution Optimization Over Multi Cloud Storage," J. Theor. Appl. Inf. Technol., vol. 100, no. 5, pp. 1378–1389, 2022.

[101] H. Sallehudin, R. Che Razak, M. Ismail, A. F. Md Fadzil, and R. Baker, "Cloud Computing Implementation in The Public Sector: Factors and Impact," Asia-Pacific J. Inf. Technol. Multimed., vol. 07, no. 02(02), pp. 27–42, 2018, doi: 10.17576/apjitm-2018-0702(02)-03.