International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 7 Issue 4 Ver. VI|| April 2018 || PP 11-15

Review on Cost Effective Data Replication Techniques in a Cloud-based environment

Sheril Sophia D'couto[1], Dhina Suresh[2]
[1]*Research Scholar, Department of Computer Science,*
St. Joseph's College of Arts and Science for Women, Hosur, Tamilnadu, (India)
[2] *Assistant Professor, Department of Computer Science,*
St. Joseph's College of Arts and Science for Women, Hosur, Tamilnadu, (India)
Corresponding auther: Sheril Sophia D'couto

Abstract : *The Need For Storing Huge Amounts Of Data Has Paved The Way For Server Virtualization And Subsequently Cloud Computing. This Has, To A Large Extent, Replaced The Traditional Models Of Data Storage. The Need To Also Preserve The Integrity Of The Data That Is Stored On The Cloud Has Resulted In Huge Amount Of Research Being Conducted Towards This Field. One Such Aspect Is Data Replication, Which Allows Multiple Copies Of The Same Data To Reside In Different Geographical Locations In Such A Way That In The Event Of A Natural Disaster Or Other Unforeseen Events In A Particular Region, The Data Still Is Safe In Another Remote Region, Thereby Ensuring Integrity Of The Data Sent Over To The Cloud. The Different Techniques And Researches Aimed At Data Replication Conducted By Various Authors Are Reviewed And Analyzed In This Study.*

Keywords *-Cloud Computing, Compliance, Data Replication, Security, Virtualization*

Date of Submission: 10-04-2018 Date of acceptance: 26-04-2018

I. INTRODUCTION

The transparent and on-demand availability of computational resources that can be accessed at high speed and released with least effort made for the service provider is what is referred to as a cloud computing model. The major aim of the cloud computing systems is to ensure secure and reliable communication while handling the cost and performance of the systems. Security, for example is one factor which usually incurs a lot of cost, especially if the users are permitted unlimited access and one which should not be compromised at any cost. The major components of cloud in terms of security are Confidentiality, Integrity and Availability (CIA) of the data that is stored within the cloud. Among these three, availability ensures that the data that is stored in the cloud is easily and readily retrievable by the user as and when he wishes to retrieve it [1]. The other two aspects, confidentiality and Integrity has to be maintained along with availability to ensure that the data that is stored is not altered/modified by any means and also it also ensures that only authorized personnel are allowed to access the same. Due to the limitations of the cloud and the various issues that arise related to virtualization and multi-tenancy in a cloud-based architecture, the data available within the cloud is not protected by external factors that may tamper damage or even cause loss of data. It is very important to ensure that data is available within the cloud storage and has to be maintained within the service level agreements (SLAs) between the client and the service provider. The availability requirements differ from customer to customer and it is based on applications that are in real-time [2]. Data replication is one such technique that has been utilized since the early cloud-based storage models, even though it's only objective was fault tolerance initially. Multiple copies of the files are saved in parallel along with the original file when they are stored. Also, when the original files change their versions, the changes do also reflect to the replicas available in file systems. This has resulted in increasing the level of parallelism within the transactions. With replication [3], the amount of data that is read and written at a given time within the cloud has increased. With the increase in the size of files involved, the amount of time required during computations and file transmissions also increases which is of primary concern. However increased use of Data replication within the cloud has resulted in increased availability of the files within the cloud by substantially reducing the time taken to retrieve the files and also ensuring that data is never compromised. Replication factor has been used, to identify or control the number of copies of a file and it also informed the file system about the quantity of copies that are available for the maintenance of a particular file. The greater the availability of the file depends upon the larger the use of replication factor [4]. Therefore, this factor proportionally reduces the available memory in the system of files due to available quantities of copies required to be maintained for a file. According to traditional systems, the data within the file system is replicated by using an established replication factor. The need for parallelism is difficult to satisfy hence it is considered as

the major issue in the replication policy [5]. It is necessary for different computers to work together in the cloud for the execution of files. For this purpose, two solutions are currently in place. In the first solution, the entire file must be stored in the each one of the virtual machines or nodes available within the cloud. This by far is the best in ensuring the integrity of the data. In the second, the number of copies can be maintained below the number of virtual machines (nodes) and all the participating nodes process each request in order to save a copy of the file. There is limitation in the second solution that how the files are treated in order of gigabytes and terabytes, the reading time of the disk, the transmission files and the network traffic [6]. If the size of a file increased, it takes a toll on the processing required affecting I/O read/write capability, network, CPU, memory etc. Hence, the primary concern with use of a replication policy is that it is difficult to manage large volume of data.

II. Literature Review

Gabriel Heleno et.al (2016) presented in this paper [7] that in order to recover failures within the cloud computing scenarios, data replication technique has been applied in many cases. The performance of execution related to an application can also be enhanced with the help of data replication. In order to minimize the execution of application of bioinformatics, a policy is presented in this paper for data replication of computational cloud. A safe computing mechanism was provided through data replication in the filtering process with the application of this method.

W. Delishiya Moral, et.al (2016) proposed in this paper [8] a fragmentation and sole replication (FASR) methodology for enhancing the security and data retrieval time of cloud. The file is fragmented with this technique and each fragment is placed in a distinct node within the cloud. Each fragment is replicated only once which is what is meant by sole replication. This ensures that the data is secure within the system in which each fragment can be replicated at maximum one time. The file can be downloaded, updated as well as uploaded by the user repeatedly here. Each time the user requests the file, the required fragments are fetched and consolidated and presented to the user whereupon he can modify the required data and upload it back to the cloud. The non-cryptographic nature of the data helps in increasing the performance of the application using this methodology. Uploading of only the updated/modified fragment in the cloud node requires another automatic updating technique which is proposed here. This is a secure and time-saving mechanism provided along with the reduction in resource utilization by providing automatic updating method using the FASR technique.

MuhannadAlghamdi, et.al (2017) presented in this paper [9] the File replication problem (FRP) within the data centers. The total energy consumption while accessing the data files within the data centers is to be minimized here. The net cost reduction in resources utilized by the file replication is provided with this algorithm. Two algorithms namely, energy and time efficient heuristic file replication algorithms are presented in this paper as well. The various algorithms are compared in this paper in order to provide evaluations of performance of proposed algorithm. The two varieties of network parameters are utilized in order to evaluate performance of approximation algorithm which shows that this algorithm performs better than other algorithms.

DejeneBoru, et.al (2015) presented in this paper [10] the energy consumption and bandwidth related models in order to provide data/database access within the cloud computing datacenter. An energy efficient replication strategy is proposed in this paper which is based on the existing models. The major objective here is on Quality of Service (QoS) of these systems along with the reduction in delays during communication. The performance and energy efficiency tradeoffs are studied in detail as per the simulation results achieved in this paper. The energy consumption, utilization of bandwidth and delays in communication are minimized as per the results observed.

Moise W. Convolbo, et.al (2016) proposed in this paper [11] a job scheduling algorithm known as DRASH (Data-Replication Aware Scheduler). The data that is being transferred is secured and the unnecessary data replication is prevented with the help of this proposed technique which helps in enhancing the overall performance of systems by enforcing data locality. On the basis of hosted data, the data centers are prioritized by the proposed technique. Further, the loads of individual jobs are balanced in order to reduce the make span. In order to perform tasks, a water-filling model of an algorithm is proposed here. Through the simulations performed and results achieved it is seen that in terms of job completion time and data replication factors, DRASH performs better in comparison to other techniques.

Ismaeel Al Ridhawi, et.al (2015) presented in this paper [12] that there is a need of reliable and continuous access to data from the cloud service providers in order to handle the increasing number of mobile devices that use cloud. A location prediction technique is proposed in order to support the accessing of data in continuous manner such that the high user density locations can be identified here. The user requested data is replicated exclusively in the third party cloud services with the help of partial data replication algorithm. The data access time is reduced in the proposed technique which works on the location of cloud server and bandwidth provided. In order to evaluate the performance of proposed algorithm, the simulations are performed

and the results show that there in order to access data and avoid replication, the proposed algorithm provides better results.

IrfaanCoonjah, et.al (2016) presented in this paper [13] that when the various sizes of packets are transmitted the reliability of UDP tunnel is required along with the measurement of packet drops. The UDP tunnel is preferred in cases where the packets of various sizes are to be transmitted. This study is presented in order to determine whether or not UPD tunnel can be utilized as the mode of transmission. In order to analyze the performance of this algorithm, simulation tests are performed. As per the simulation results, it is seen that the UDP tunnel can be utilized as a reliable mode of transfer in the data centers of cloud.

SarraSlimani, et.al (2017) presented a study related to the major service replication strategies applied in cloud computing. The service dependencies of almost all the proposed methods are not provided. A novel service replication mechanism known as Dependency aware Dynamic Service replication strategy (DDSoR) is proposed in this paper further [14]. Within the service oriented applications, the reduction of execution time is the major objective of DDSoR approach. In order to achieve this objective, the correlated services are consolidated within the same server with the help of replication. On the basis of a spectral clustering, the service dependencies are identified and then the service clusters are co-located in this approach. The improvement in performance is shown as per the simulation results achieved after conducting experiments.

Wan-Chi Chang, et.al (2017) proposed in this paper an adaptive replication technique for MOS (Mobile Online System) by utilizing MECs (Mobile-edge Cloud Computing) within the 5G network. The number of masters as well as slaves will be determined in periodic manner with the help of MECs [15]. They will utilize the read/write frequency in order to do so. The response time and traffic overhead are minimized through the application of proposed technique as per the simulation results. So, from the MECs, MOS with rigid QoS can be benefitted. It is seen from the experimental results that the network traffic as well as the costs of storage are affected by the read/write ratio of operations.

AnandTripathi et.al (2016) proposed in this paper a scalable protocol which can provide transactions with remote partitions writes within the cloud computing systems that cause partial data replication [16]. The notion of escrow based ordering of remote partition updates is utilized in order to improve the scalability of the PCSI model through this protocol. Along with this, it ensures the availability of causal consistency within these systems. The issues related to the escrow mechanism are presented within this paper as well. By utilizing three various workload classes as well as five benchmarks on the cluster which includes 100 nodes, the scalability of the proposed protocol is evaluated in this paper. As per the results it is seen that the transactions within the partially replicated data systems are provided a scalable approach through this proposed technique.

Nikos Tziritas, et.al (2016) presented in this paper a study related to the utilization of replication within the conjunction along with the issue related to VM assignment. Amongst the already existing cloud and mobile cloud systems, it is to be studied as to where and which VM migration is to be performed such that the network overhead can be minimized [17]. On the basis of hyper-graph partitioning, an algorithm is proposed in this paper such that the issues which arouse previously can be resolved. The capture storage as well as computing capacity constraints related problem is to be solved through this proposed algorithm. There is minimization of around 53% of network overhead as per the experimental results achieved after performing simulation evaluations and it is concluded that the proposed algorithm has outperformed other existing approaches.

TABLE 1: A SURVEY ON EXISTING DATA REPLICATION TECHNIQUES

Author's Names	Year	Description	Outcome
Gabriel HelenoGonçalves da Silva, MaristelaHolanda-e-AleteiaAraujo	2016	In this paper it is presented that in order to recover failures within the cloud computing scenarios, data replication technique is applied in many cases.	There is no need for reruns here. The file system data can be accessed here as well.
W.Delishiya Moral, B.Muthu Kumar	2016	In this paper a fragmentation and sole replication (FASR) is presented for enhancing the security and data retrieval time of cloud.	There is secure and time-saving mechanism provided along with the reduction in resource utilization by providing automatic updating method FASR technique.
MuhannadAlghamdi, Bin Tang, and Yutian Chen	2017	In this paper the file replication problem (FRP) within the data centers is presented.	The two varieties of network parameters are utilized in order to evaluate performance of approximation algorithm which shows that this algorithm performs better than other algorithms.
DejeneBoru, DzmitryKliazovich, FabrizioGranelli, Pascal Bouvry, Albert Y. Zomaya	2015	An energy efficient replication strategy is proposed in this paper which is based on the existing models.	The energy consumption, utilization of bandwidth and delays in communication are minimized as per the results.
Moise W. Convolbo, Jerry Chou, Shihyu Lu and YehChing	2016	In this paper a job scheduling algorithm known as DRASH (Data-Replication Aware Scheduler) is	Through the simulations performed and results achieved it is seen that in terms of job completion time and data replication factors, DRASH

Chung		proposed.	performs better in comparison to other techniques.
Ismaeel Al Ridhawi, NourMostafa, WassimMasri	2015	A location prediction technique is proposed in order to support the accessing of data in continuous manner such that the high user density locations can be identified here.	In order to evaluate the performance of proposed algorithm, the simulations are performed and the results show that there in order to access data and avoid replication, the proposed algorithm provides better results.
IrfaanCoonjah, Pierre Clarel Catherine, K. M. S. Soyjaudah	2016	In this paper it is presented that when the various sizes of packets are transmitted the reliability of UDP tunnel is required along with the measurement of packet drops.	As per the simulation results, it is seen that the UDP tunnel can be utilized as a mode of transfer in the data centers of cloud.
SarraSlimani, TarekHamrouni, Faouzi Ben Charrada, Frederic Magoules	2017	A novel service replication mechanism known as DDSoR is proposed in this paper.	The improvement in performance is shown as per the simulation results achieved after conducting experiments.
Wan-Chi Chang, Pi-Chung Wang	2017	In this paper an adaptive replication technique is proposed for MOS by utilizing MECs within the 5G network.	It is seen from the experimental results that the network traffic as well as the costs of storage are affected by the read/write ratio of operations.
AnandTripathi and GowthamRajappan	2016	In this paper a scalable protocol is proposed which can provide transactions with remote partitions writes within the cloud computing systems that cause partial data replication	As per the results it is seen that the transactions within the partially replicated data systems are provided a scalable approach through this proposed technique.
Nikos Tziritas, Maria Koziri, AretiBachtsevani, ThanasisLoukopoulos, George Stamoulis, Samee U. Khan	2016	In this paper a study related to the utilization of replication is presented within the conjunction along with the issue related to VM assignment.	There is minimization of around 53% of network overhead as per the experimental results achieved after performing simulation evaluations and it is concluded that the proposed algorithm has outperformed other existing approaches.

III. CONCLUSION

This paper reviews the various techniques used in data replication in a cloud computing environment. Regulatory compliance laws across countries require organizations like banks or financial institutions to preserve their electronic data for a predefined amount of time without compromising on the confidentiality, integrity and its availability. It has been increasingly challenging for these organizations to store the data, maintaining the above three aspects without investing a huge fortune on the same. The cloud solution with its on-demand based resource allocation seems to be the way ahead. Not much research has been done so far in this field and Banks/Financial institutions currently rely on traditional model of storing data.The analysis of each technique has been examined and has given a bird's eye view of the different methodologies used so far and has helped in visualizing a solution that would be best suitable for a compliance driven cloud model**.**

REFERENCES

[1] FarzadSabahi, "Cloud Computing Reliability, Availability and Serviceability (RAS): Issues and Challenges", 2011, International Journal on Advances in ICT for Emerging Regions, Volume: 04 (02) : 12 – 23.

[2] Andres J. Gonzalez, Bjarne E. Helvik, PrakritiTiwari, Denis M. Becker and Otto J. Wittner, "GEARSHIFT: Guaranteeing Availability Requirements in SLAs using Hybrid Fault Tolerance", 2015, Proc. IEEE Conference on Computer Communications (INFOCOM).

[3] Mohamed-K HUSSEIN, Mohamed-H MOUSA, "A Light-weight Data Replication for Cloud Data Centers Environment", 2014, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 1.

[4] DejeneBoru, DzmitryKliazovich, FabrizioGranelli, Pascal Bouvry, Albert Y. Zomaya, "Energy-Efficient Data Replication in Cloud Computing Datacenters", Globecom 2013 Workshop - Cloud Computing Systems, Networks, and Applications.

[5] SagarVerma, Arun Kumar Yadav, Deepak Motwani, "An Overview of Data Replication and Load Balancing Issues in Cloud Computing Environment", 2015, IJITKM, Volume-8, Number-2, pp. 71-76.

[6] Joarder Mohammad Mustafa Kamal and ManzurMurshed, "Distributed Database Management Systems: Architectural Design Choices for the Cloud", 2014, Springer International Publishing Switzerland.

[7] Gabriel HelenoGonçalves da Silva, MaristelaHolanda-e-AleteiaAraujo, "Data Replication Policy in a Cloud Computing Environment", 2016, IEEE.

[8] W.Delishiya Moral, B.Muthu Kumar, "Improve the Data Retrieval Time and Security through Fragmentation and Replication in the Cloud", 2016, Proc. International Conference on Advanced Communication Control and Computing Technologies (ICACCCT).

[9] MuhannadAlghamdi, Bin Tang, and Yutian Chen, "Profit-Based File Replication in Data Intensive Cloud Data Centers", IEEE ICC 2017 SAC Symposium Cloud Communications and Networking Track.

[10] DejeneBoru, DzmitryKliazovich, FabrizioGranelli, Pascal Bouvry, Albert Y. Zomaya, "Models for Efficient Data Replication in Cloud Computing Datacenters", IEEE ICC 2015 - Communications QoS, Reliability and Modeling Symposium.

[11] Moise W. Convolbo, Jerry Chou, Shihyu Lu and YehChing Chung, "DRASH: A Data Replication-Aware Scheduler in Geo-distributed Data Centers", 2016 IEEE 8th International Conference on Cloud Computing Technology and Science.

[12] Ismaeel Al Ridhawi, NourMostafa, WassimMasri, "Location-Aware Data Replication in Cloud Computing Systems", 2015 Eight International Workshop on Selected Topics in Mobile and Wireless Computing.

[13] IrfaanCoonjah, Pierre Clarel Catherine, K. M. S. Soyjaudah, "Evaluation of UDP Tunnel for Data Replication in Data Centers and Cloud Environment", Proc. International Conference on Computing, Communication and Automation (ICCCA2016).

[14] [14] SarraSlimani, TarekHamrouni, Faouzi Ben Charrada, Frederic Magoules, "DDSoR: A Dependency aware Dynamic Service Replication strategy for efficient execution of service-oriented applications in the cloud", Proc. 2017 International Conference on High Performance Computing & Simulation.

[15] [15] Wan-Chi Chang, Pi-Chung Wang, "An Adaptable Replication Scheme in Mobile Online System for Mobile-edge Cloud Computing",Proc, 2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS): IECCO: Integrating Edge Computing, Caching, and Offloading in Next Generation Networks.

[16] [16] AnandTripathi and GowthamRajappan, "Scalable Transaction Management for Partially Replicated Data in Cloud Computing Environments", Proc, 2016 IEEE 9th International Conference on Cloud Computing.

[17] [17] Nikos Tziritas, Maria Koziri, AretiBachtsevani, ThanasisLoukopoulos, George Stamoulis, Samee U. Khan, "Data Replication and Virtual Machine Migrations to Mitigate Network Overhead in Edge Computing Systems", 2016, IEEE.

Sheril Sophia D'couto "Review on Cost Effective Data Replication Techniques in Cloud-based environment ""International Journal of Engineering Science Invention (IJESI), vol. 07, no. 04, 2018, pp 11-15