iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 6 **Issue:** V **Month of publication:** May 2018

**DOI:** http://doi.org/10.22214/ijraset.2018.5211

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887*
*Volume 6 Issue V, May 2018- Available at www.ijraset.com*

# An Approach for Replica Creation & Replica Placement using MCDM

Amit Kumar Singh[1], Udai Shanker[2]
[1, 2]*Department of Computer Science & Engineering Madan Mohan Malaviya University of Technology Gorakhpur, India*

***Abstract:*** ***Due to huge amount of data creation day by day in all the fields, maintaining and managing the data in effective manner is an arduous task. To increasing the availability of data, distributed data storage system is playing a very essential role. Huge amount of data is created for many purpose and handling of this data, data management is required. To improve the performance of the system in which data is not locally available & data is shared to many users to increase the performance of the system, data replication is the technique which is used to improve the performance of availability & decrease the latency of system. Maintaining the consistency of data in replicated data system is today's burning topic of research. Very few methods have been proposed for maintaining the consistency while replica creation and replica placement. In this paper, a combination of replica creation & replica placement algorithm has been proposed in which most popular file is obtained using the calculation of access frequency of all the data items & replica placement is performed using multi criteria decision making technique. analytical hierarchy process is used for the selection of best node for replica placement considering various criteria for selection. The simulation results show that the above scheme outperforms than other techniques***

***Keywords: Multi Criteria Decision Making, Analytical Hierarchy Process, Replica Creation, Replica Placement, System Performance, Availability, Database Systems***

## I. INTRODUCTION

Cloud data storage become popular because of handling huge amount of data is required & size of data is increasing day by day. In cloud data storage system replication is used for creating similar copies of some data items stored in the cloud. Latency can be minimised after creating replication at various data centres from source data centres. Response time is also crucial factor these days so replica creation & its placement to right place is important for the system performance & decrease the latency of the application deployed over the cloud. Every user wants to use only that replica which is geographically nearby its location to response time can be minimized. For continuous data, a system must be designed with no failure which is practically impossible, so we can create replica of data so that if some of the replica is not able to serve the request another replica can process the request & application execution complete successfully. We cannot place replica at all the nodes because every node has some limitation some of them have limited storage & some of them can be minimum bandwidth, so we must find out a new technique for replica placement. The two-major problem which is minimized by data replication in cloud storage system in data availability & performance of the application. For a file, how many number of replica should be created is the first issue in this type of situation & then where to place those replicas is second challenge. Which files should be replicate, this can be obtained by the analysis of user request. Replication can be categorized in two ways known as static replication & dynamic replication. In static replication, which data is replicated at which data centre is decided in advance but in case of dynamic data replication technique no. of replica & placement of replica is decided according to the situation of the running application. In dynamic replication technique replica is created & placed at appropriate place according to the demand frequency of data items. We are using multi criteria decision method for the replica placement then we are considering following criteria such as user request, storage, bandwidth, system performance etc. here AHP multi criteria decision making is applied to solve the problem of replica placement. AHP is a mathematical method which is used for decision making. The criteria on which we are applying the AHP technique is, CPU(C), Storage (S), Bandwidth (B), No. of Request raised (R), Location (L).

## II. REPLICATION CONCEPTS

Replication is the process of sharing information between different users and to maintain consistency at the same time. Redundant resources share hardware component, software component to improve the reliability, system fault tolerance and data accessibility. Data replication is performed in the environment when the network of database and distributed system. database replication can be termed as creating and maintaining the duplicate copies of data items in distributed database system [3]. Replication is applied on

frequently used data items and data item located at one computer or server copied in another database may be placed on same computer, so the user can access the information with more reliable and with high speed [4]. Database replication provide option to choose data items from distinct locations to improve data availability, improving performance of the system and overcome the situation of system failure of applications. If user trying to access the data item from a source and high traffic is available on that path, then user can access the same data item from different location. The overall objective of data replication is to improve the scalability of the distributed system. Data Replication is performed for achieving some objective in the system [4].

### *A. Database Locality*

A replicated distributed database system maintains the database locality. Every user wants to access the data from that source which is nearby because the communication speed is high as compare to other sources which are geographically at a long distance from the user. Providing closest database for user increase the performance of entire system.

### *B. Performance*

Data Replication in distributed environment improve both the database operation read operation and write operation. Dealing with read operation and write operation simultaneously is arduous task. When database is stored at a single server and used by so many users over the network the performance of the database system may decreased if so many no. of request raised at the same time. Multiple copies of data offer users to access data items from distinct locations which improve the performance.

### *C. Availability And Fault Tolerance*

Replication is mainly implemented in database system to increase the availability of data. In database system downtimes occurs when database system not able to respond. Sometimes downtimes are created when maintenance is required on database system & sometimes downtime occurs because of some unpredictable reasons like software failure, hardware failure, human error etc. In case of downtime database system not able to provide services to its user but if replication is applied on database system then data can be obtained from another server if one server is down. If data is stored at only a single server, the availability of database is compromised. Replica is also used for restoring the database system in case of server failure, so we can understand that how the replication is applied to increase the availability of data [4].

## III. LITERATURE REVIEW

Some works has been done in the field of cloud-based storage system and distributed database system to increase the overall performance of the system. Some work has been done in field of replication to improve the system performance.

### *A. Static Replica Creation Strategy*

Static replica creation technique is first applied in Hadoop distributed file system in which the no. of replica is defined from the client side & if the no. of replica is not given the by default 3 replicas is crated. Two of them placed to the neighbor of client node & on is randomly placed to other nodes. If more than three replicas need to place, then replica are assigned randomly. Static replica creation reduces the latency & increase data availability, but static replication cannot be applied in all the situation. When the static replica creation is applied then storage capacity of the nodes is not considered so it can create problem in some case when there is no space for stopping criteria.

### *B. Dynamic Replica Creation Strategy*

Dynamic replication technique is used for flexible availability of data replication. According to the access frequency of the data items no. of replica can be generated and replica can be placed to the appropriate place. Dynamic replica creation technique improves the performance of the system. In static replica creation replica are created in advance & no. of replica required or not but the replica have to be created so unnecessary storage is acquired. Dynamic replication can be applied considering the following criteria to Minimize Risk, minimize expected utility, Best Client, minimizing time, Maximized performances, Time utility and so on. Replica are created to those nodes where the expected utility is lowest. For Each criterion, the replica placement technique is different.

Literature [1] introduces six methods of replica creation and placement technique, are: no replica strategy, caching strategy, the Best Client strategy, waterfall strategy, waterfall combining with caching strategy, and rapid spread strategy. The caching strategy and Best Client strategy creates a replica based on no of user visited. This technique is not enough good to maintain the decorum of replicated system. All the above strategies suggested are not enough to handle the amount of data stored in Cloud data storage.

Vickers–Clarke–Groves presented a technique for replica placement using considering the concept of acceleration [3]. Based on accessing locality of files within the system replica of the files is generated and replica is placed to that node which request more for some data items. The method of replica creation proposed by Vickers Clark groves is also known as VCG mechanism.

Geo-cloud based two-layer dynamic replica creation strategy called TG stag is proposed [5]. TG stag addresses the issue with twofold strategy: policy constraint heuristic inter-datacenter replication and load aware adaptive intra-datacenter replication.TG stag aims to reduce the cross-data center bandwidth and access time consumption

TOPSIS method is also used for dynamic replica creation in cloud storage system. The nodes in the network is sorted by the TOPSIS method and as per the performance of the node and load on the nodes replica are placed to the nodes.

Raja Lakshmi et al (2014) propose a method for dynamic data replication in cloud. The proposed approach applied on cloud storage for replica creation and replica placement to maintain the availability of data items within the system [6]. The method consists of two main phases: file application and replication operation. The first phase deal with replica creation and location of the replicated by using catalog and index, whereas the second phase is applied to perform optimization among replicated copies.

## IV. MULTI CRITERIA DECISION MAKING

Multi criteria decision making is a technique used to make decision of selecting the appropriate solution if the set of solutions exist. In multi criteria decision making technique the solution is obtained considering the various criteria for getting the effective solution. If we have a problem & many solution exist & we are looking for the solution considering some criteria, then different multi criteria decision making technique can be a useful technique. Multi criteria decision making used to select the best alternative from the pre-selected solutions. The final decision in multi criteria decision making obtained considering various criteria.

A decision-making process passes through the following steps for making decision

*A.* what is the objective of the decision making
*B.* what are the criteria for making decision.
*C.* are the alternatives available.
*D.* what are the parameters to choose the relative order of criteria.
*E.* Aggregation is performed based on importance factor of criteria.
*F.* Decision is based on aggregation result.

Analytical Hierarchy Process is one of the most popular method of MCDM which is used to make decision from different available alternatives. In AHP method the criteria are measured on different scale & it may possible that those units are having different units. In AHP method pairwise comparison is performed & result is mainly dependent on user perception. AHP is applied only in that case where finite no. of alternatives available [9]. Decision is based on some restriction. AHP process consist of following steps:

Step1: Collection of Information.
Step2: Quantification of the Information.
Step3: Modeling
Step4: Optimization Process.

AHP process based on some user perception because the quantification is mainly based on some user perception, so we must check the consistency ratio of the eigen values obtained during the process of AHP. So, if the consistency ration is less than a threshold value then only accepted else the decision makers must consider some other datasets or to make decision with some different perception.

## V. PROPOSED SOLUTION

Managing huge amount of data and maintain the availability & increasing the performance of the system where so many nodes are connected to provide parallel solution data replication is important technique. The main problem of replicated system is to maintain the consistency of data and how to manage the data up to date within the entire system. Initially so many algorithms proposed to solve the consistency problem & maintaining the data up to date. Now the problem of replica creation and replica placement come into existence. Here in our proposed solution we are calculating the most popular file within the system & with the help of request raised within the given time interval we are trying to find out most popular file within that time interval & after calculating most popular file we find how many replicas is to be created for that data item and then selection of appropriate node for replica placement. The most popular file is to be replicated to the node, so the selection of appropriate node is selected using multi criteria decision method. So, the combination of both the technique to find the most popular file & replica placement to the appropriate node

is surely improve the performance. Here we are using analytical hierarchy process to make decision about the selection of the best node.

*A. Calculation of Most Popular file*

Here we consider that four nodes are communicating to each other data items are stored at all the nodes nut not each data item is stored at all the nodes so if a node transaction needs a data item which is not store on that node then they get the data item from the other node. So, in the system a header keep the record of no. of request for a data item by particular node in the given time interval.

| Data Item | Access Request by | Access Time |
|---|---|---|
| A | N2 | 10 |
| A | N3 | 15 |
| B | N4 | 17 |
| E | N2 | 5 |

Record of Node1 Header

| Data Item | Access Request by | Access Time |
|---|---|---|
| D | N1 | 5 |
| D | N4 | 15 |
| E | N2 | 6 |

Record of Node2 Header

| Data Item | Access Request by | Access Time |
|---|---|---|
| B | N3 | 8 |
| D | N1 | 7 |

Record of Node3 Header

| Data Item | Access Request by | Access Time |
|---|---|---|
| A | N3 | 5 |
| E | N2 | 13 |

Record of Node4 Header

Now with the help of all the record we can obtained the aggregate record for data items. The aggregate record shows the total no. of access request in the given time interval.

| Aggregate Record | |
|---|---|
| Data Item | Access Time |
| A | 30 |
| B | 25 |
| D | 27 |
| E | 24 |

After collecting the record from the all the nodes, we calculate the access frequency of the data items in the given time interval. Assume *NT* is the number of time intervals passed, A is the set of files that have been requested, and $a^t_d$ indicates the number of accesses for file *f* at time interval *t*. The *AF* for file *f* is calculated as:

$$AF(d) = \sum_{t=1}^{N_T} (a^t_d * 2^{-(N_T - t)})$$

After calculating the Access Frequency of each data item in the given time interval the data item with maximum access frequency is known as most popular file within that time interval. Now we must find out the no. of replica is to be created to increase the availability within the system of most popular file.

$$\text{No. of Replica} = \frac{\text{Average Frequency of Most Popular File}}{\text{Average Frequency of All the file}}$$

*B. Replica Placement Using AHP Method*

To place the replica, we must consider three thing objectives (Replica Placement) Criteria (Which factor decide where to place replica) alternatives (What are the options available to place the replica).

1) *Step1:* A Hierarchy for Replica Placement using Multi Attribute Decision Making is created….

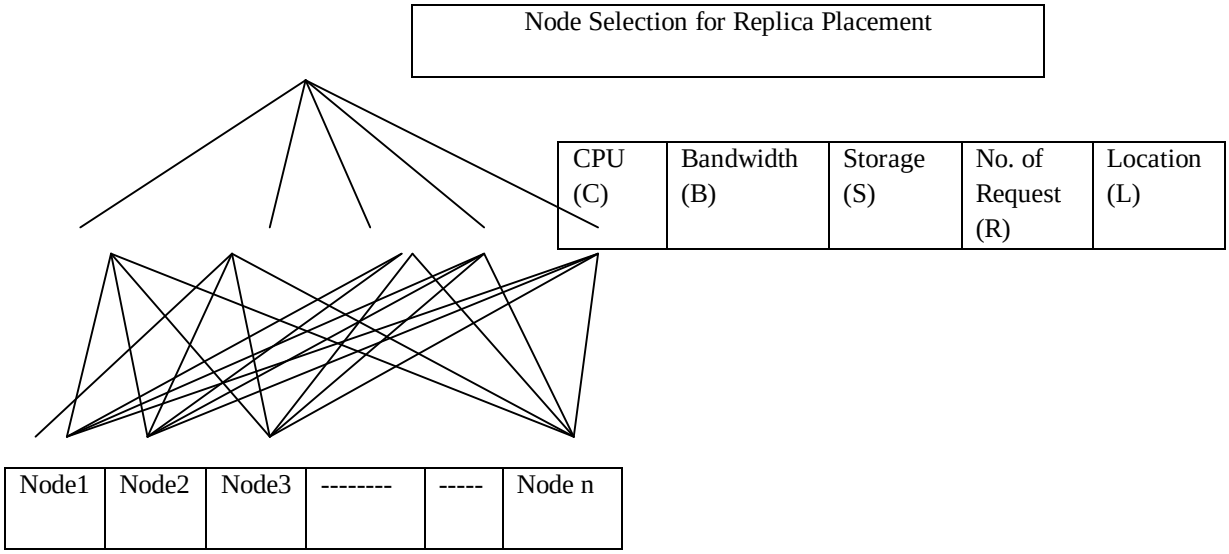



Figure 1 Node Selection based on Criteria

2) *Step2:* Create a Two-Dimensional Matrix of Pairwise Comparison of Criteria.

| | C | B | S | R | L |
|---|---|---|---|---|---|
| C | 1 | C/B | C/S | C/R | C/L |
| B | B/C | 1 | B/S | B/R | B/L |
| S | S/C | S/B | 1 | S/R | S/L |
| R | R/C | R/B | R/S | 1 | R/L |
| L | L/C | L/B | L/S | L/R | 1 |

Table I Relationship Among Criteria

3) *Step3:* Similarly, we obtained the pairwise matrix for all other alternatives & its Criteria So,

For C:

$$\begin{pmatrix} C_{11} & C_{12} & C_{13}\ldots\ldots\ldots\ldots C_{1n} \\ C_{21} & C_{22} & C_{23}\ldots\ldots\ldots\ldots C_{2n} \\ & & . \\ & & . \\ C_{n1} & C_{n2} & C_{n3}\ldots\ldots\ldots\ldots C_{nn} \end{pmatrix}$$

Table II CPU Relations

For B:

$$\begin{pmatrix} B_{11} & B_{12} & B_{13}\ldots\ldots\ldots\ldots B_{1n} \\ B_{21} & B_{22} & B_{23}\ldots\ldots\ldots\ldots B_{2n} \\ & & .. \\ & & ---- \\ B_{n1} & B_{n2} & B_{n3}\ldots\ldots\ldots\ldots Bnn \end{pmatrix}$$

Table III: Bandwidth Relations

For S:

$$\begin{pmatrix} S_{11} & S_{12} & S_{13}\ldots\ldots\ldots\ldots S_{1n} \\ S_{2} & S_{22} & S_{23}\ldots\ldots\ldots\ldots S_{2n} \\ & & . \\ & & . \\ S_{n1} & S_{n2} & S_{n3}\ldots\ldots\ldots\ldots Snn \end{pmatrix}$$

Table IV Relationship Among Storage

For R:

$$\begin{pmatrix} R_{11} & R_{12} & R_{13}\ldots\ldots\ldots\ldots R_{1n} \\ R_{21} & R_{22} & R_{23}\ldots\ldots\ldots\ldots R_{2n} \\ & & . \\ Rn_{1} & R_{n2} & R_{n3}\ldots\ldots\ldots\ldots Rnn \end{pmatrix}$$

Table V Relationship Among No. of Request

For L:

$$\begin{pmatrix} L_{11} & L_{12} & L_{13}\ldots\ldots\ldots\ldots L_{1n} \\ L_{21} & L_{22} & L_{23}\ldots\ldots\ldots\ldots L_{2n} \\ & & . \\ & & . \\ L_{n1} & L_{n2} & L_{n3}\ldots\ldots\ldots\ldots Lnn \end{pmatrix}$$

Table VI Relationship Among Location of the Node

*4) Step4:* when pairwise matrix is obtained then we should create a priority matrix for all the matrix above explained. Here we calculate eigen values of all the matrix

Calculation of eigen vector of matrix in step 2 & Relative priority is obtained as

$$\begin{pmatrix} RPC \\ RPB \\ RPS \\ RPR \\ RPL \end{pmatrix} \quad \ldots\ldots\ldots \text{Eqn A}$$

Table VII Relative priority of criteria

Similarly, we obtain the eigen values of all the matrix in step3 & calculate the relative priority.

$$\begin{pmatrix} RPC_1 & RPB_1 & RPS_1 & RPR_1 & RPL_1 \\ RPC_2 & RPB_2 & RPS_2 & RPR_2 & RPL_2 \\ & \ldots & \ldots & \ldots & .. \\ RPC_n & RPB_n & RPS_n & RPR_n & RPL_n \end{pmatrix} \quad .. \text{Eqn B}$$

Table VIII Relative Priority of all the nodes Criteria

5) *Step 5:* Now from the equation A and Equation B we can obtain the Relative weight of nodes:

| | |
|---|---|
| Node1 | RPC*RPC1 + RPB*RPB1 + RPR*RPR1… |
| Node2 | RPC*RPC2 + RBB*RPB2 + RPR*RPR2… |
| Node3 | RPC*RPC3 + RPB*RPB3+ RPR*RPR3…. |
| Node4 | RPC*RPC4 + RPB*RPB4 + RPR*RPR4…. |
| . | |
| . | |
| Node n | RPC*RPCn+ RPB*RPBn+ RPR*RPRn... |

Table IX Node Selection Based on

## VI. CONCLUSION

Data replication is easy solution when data is highly distributed & we must keep care about database system performance & availability. Replica are created to increase the performance of the system in which data items are distributed to the users. Data Replication increase the accessibility of the shared database system in which all the replica synchronized to each other. Dynamic data replication helps in such cases like storage & overloading of the system. In this paper, a combined method of replica creation and replica placement in dynamic environment is proposed. Replica placement is performed after analysis of various nodes in the distributed system so after the implementation of this technique we can see the improvement in the response time of the system.

## REFERENCES

[1]. Foster and K. Ranganathan, “Design and Evaluation of Dynamic Replication Strategies for a High-Performance Data Grid,” … Int. Conf. Comput. High …, 2001.

[2]. R. M. Rahman, K. Barker, and R. Alhajj, “Replica placement in data grid: considering utility and risk,” Int. Conf. Inf. Technol. Coding Comput., vol. 1, pp. 354–359, 2005.

[3]. W. Hongxia, “Application of VCG in Replica Placement Strategy of Cloud Storage,” Int. J. Grid Distrib. Comput., vol. 9, no. 4, pp. 27–40, 2016.

[4]. Chang, Ruay-Shiung, and Hui-Ping Chang. "A dynamic data replication strategy using access-weights in data grids." The Journal of Supercomputing 45, no. 3 (2008): 277-295.

[5]. Acharya, Swarup, and Stanley B. Zdonik. "An efficient scheme for dynamic data replication." (1993).

[6]. Huang, Y. and Wolfson, O., 1993, April. A competitive dynamic data replication algorithm. In Data Engineering, 1993. Proceedings. Ninth International Conference on (pp. 310-317). IEEE.

[7]. Saadat, Nazanin, and Amir Masoud Rahmani. "PDDRA: A new pre-fetching based dynamic data replication algorithm in data grids." Future Generation Computer Systems 28, no. 4 (2012): 666-681.

[8]. Nicholson, C., Cameron, D.G., Doyle, A.T., Millar, A.P. and Stockinger, K., 2008. Dynamic data replication in lcg 2008. Concurrency and Computation: Practice and Experience, 20(11), pp.1259-1271.

[9]. Lamehamedi, Houda, Boleslaw Szymanski, Zujun Shentu, and Ewa Deelman. "Data replication strategies in grid environments." In Algorithms and Architectures for Parallel Processing, 2002. Proceedings. Fifth International Conference on, pp. 378-383. IEEE, 2002.

[10]. Chang, Ruay-Shiung, Hui-Ping Chang, and Yun-Ting Wang. "A dynamic weighted data replication strategy in data grids." In Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on, pp. 414-421. IEEE, 2008.

[11]. Bsoul, Mohammad, Ahmad Al-Khasawneh, Yousef Kilani, and Ibrahim Obeidat. "A threshold-based dynamic data replication strategy." The Journal of Supercomputing 60, no. 3 (2012): 301-310.

[12]. Daniell, Thomas P., Robert C. Harding Jr, Neil J. Lewis, and Sven HH Nauckhoff. "Method for the dynamic replication of data under distributed system control to control utilization of resources in a multiprocessing, distributed data base system." U.S. Patent 4,432,057, issued February 14, 1984.

[13]. Wolfson, Ouri, and Sushil Jajodia. "Distributed algorithms for dynamic replication of data." In Proceedings of the eleventh ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems, pp. 149-163. ACM, 1992.

[14]. H. Mengxing, Y. Xianglong, W. Sanpeng, and Z. Donghai, “A Strategy of Dynamic Replica Creation in Cloud Storage,” Proc. 1st Int. Work. Cloud Comput. Inf. Secur., no. Ccis, pp. 389–392, 2013.

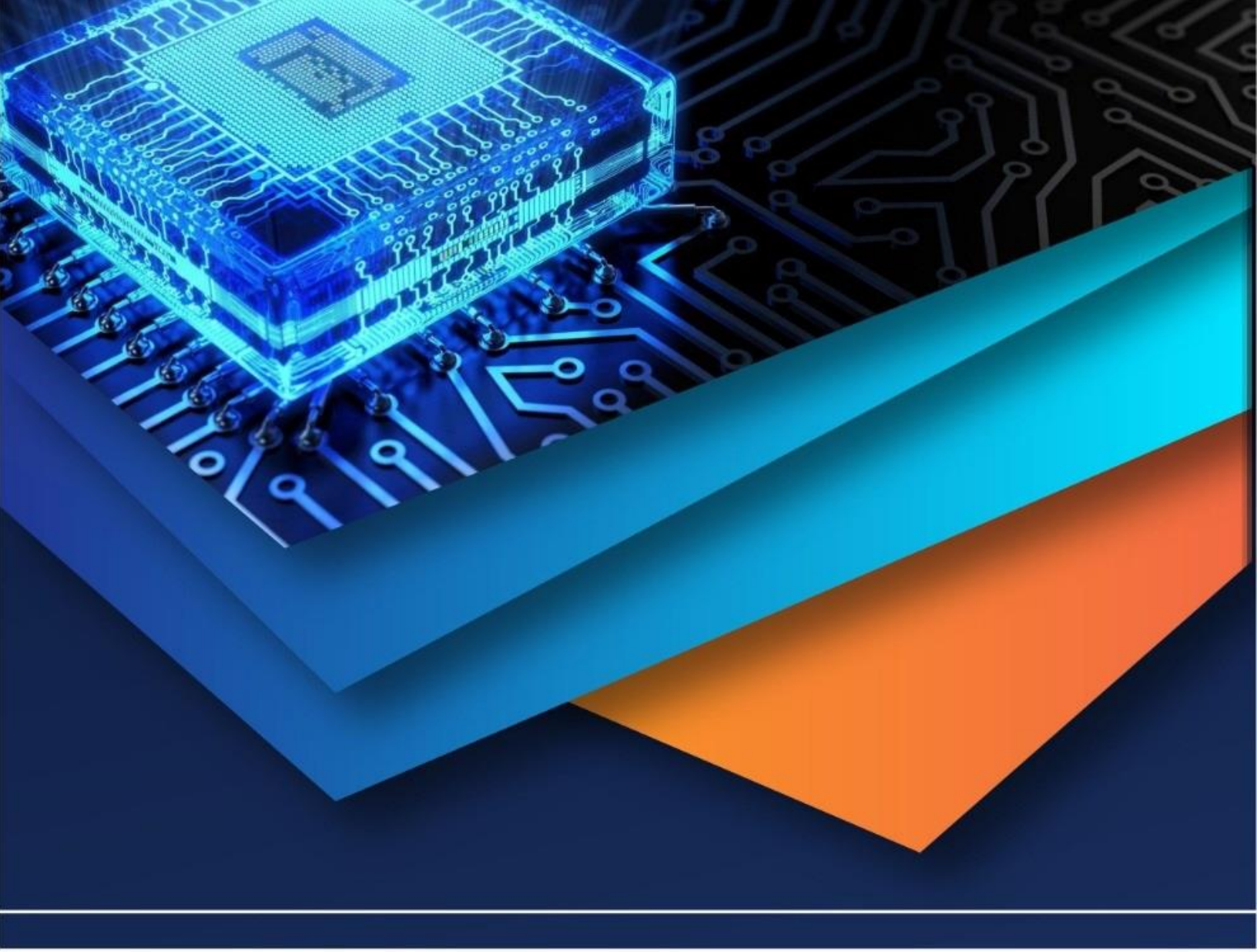
[15]. Z. Ye, S. Li, and J. Zhou, “A two-layer geo-cloud based dynamic replica creation strategy,” Appl. Math. Inf. Sci., vol. 8, no. 1, pp. 431–440, 2014.

[16]. Rajalakshmi, D. Vijayakumar, and K. G. Srinivasagan, “An improved dynamic data replica selection and placement in cloud,” 2014 Int. Conf. Recent Trends Inf. Technol. ICRTIT 2014, vol. 3, no. 3, 2014.

[17]. Salman Abdul Moiz, Sailaja P., Venkataswamy G., Supriya N. Pal. “Database Replication: A Survey of Open Source and Commercial Tools”. International Journal of Computer Applications(0975 – 8887) Volume 13– No.6, 2011

[18]. Heinz Stockinger “Data Replication in Distributed Database Systems”, 1999.

[19]. Satty, Thomas L. "Decision making—the analytic hierarchy and network processes (AHP/ANP)." Journal of systems science and systems engineering 13, no. 1 (2004): 1-35.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY