Different Data Replication Strategies in Cloud Environment

Guilherme Ferreira Gomes
Assistant Mechanical Engineering. Federal Universiry Of Itajubá, Itajubá, Brazil

ABSTRACT
Replication offers high performance, high data availability as well as it enhances the systems reliability. When the data files replica is increased, the maintenance as well as the production cost also goes beyond the limit. More consequently replication technique undergoes on three main limitations. Here, we have performed a comparative study of the different data replication strategies such as ADRS (Adaptive Data Replication Strategy), DCR2S (Dynamic Cost Aware Re-replication and Rebalancing Strategy) and EPA (Efficient Placement Algorithm) in cloud environment. The implementation of these three techniques is done in JAVA and the performance analysis is conducted to study the performance of those replication techniques by using various parameters. The parameters utilized for the comparison of performance of these three techniques are Load variance, Response time, Probability of file availability, SBER (System Byte Effective Rate), Latency, and Fault Ratio. From the analysis, it is evaluated that by varying the number of file replicas it shows deviations on the outcomes of these parameters.

Keywords: Data Replication, Data placement

I. INTRODUCTION
Cloud processing empowers the administration of tremendous IT administrations that is being developed on the topmost of topographically distributed domains and shared it entirely [1-10]. Thus, for better execution and for better reliability, the assets (resources) ought to be replicated over redundant positions as well as it utilizes the redundant frameworks [11-16]. To deploy the data traffics exponential increment and to optimize the datacenter frameworks bandwidth and energy a few data replication techniques is established. Handling the replicas over numerous sites obviously increases the execution thus by diminishing the current access delay and by the elimination of unique failure point [17-26]. In order to maintain the replicas of data, few frameworks, namely the networking devices and storage devices are being demanded. Beyond this, it is necessary to synchronize the currently generated new replicas and any of the progressions created over one of the locales (sites) required to be copied onto another position [27-35].

Data center framework intakes more energy and thus it turns out to be non-utilized. The utilized assets (resources) get to be depleted without including the extra expense. Thus, by the availability of flexible services together with better performance and tremendous data availability turns to be a necessary demand that ought to be met in different situations. In order to face this demand, the term replication is utilized [36]. Data replication is the method of maintaining data files multiple copies over multiple sites thus to enhance better availability, reliability, and performance. Replication is basically utilized in intensive data applications in which the data ought to be shared and required to be gained from distinctive sites, placed over various topographical domains. Replication builds accessibility of the data [37]. If a portion of the sites capturing the information about the file failure at the time of the request, the request will be provided from any other sites handling the replicas (copies) of the specific file [38]. Beyond this, the replication technique also enhances the entire system performance where the request from users is possible to be satisfied over closest sites maintaining the demanded data files replicas. This process maximizes the system's response time as well as the access time. Finally, the replication will also increase the reliability of the system. By chance the corruption of any of the data files replica, or else it fails because of few criteria the request from users can be provided over the different site which is being non-corrupted that maintains the specific replicas [39-46].

II. SYSTEM MODEL
The illustrations in Figure.1 depicts the multi-level hierarchical, heterogeneous cloud framework comprises of various virtual machines, brokers, host, replica catalogs, users and replica managers. By nature the data centers are basically heterogeneous. Various Service Providers controls the data centers. The Service Providers mainly choose about the replication strategies, replication cost and various issues associated with the data centers configuration. Every data center is comprised with different hosts.
The host is referred to as a physical machine which is being ought to administer various virtual machines. In order to access the data file, the user transmits the demand (request) to the broker. The position of each data files replicas is maintained by replica catalog. Because of this the broker initially transmits the request to the replica catalog in order to access the list of data centers holding the data files that is being requested [47].

After the acceptance of the data centers list capturing the details about the data file that is being requested, the request is then scheduled to the neighbor data center by the broker. Hereby the replica manager is in charge for the generation of new replicas over various domains. The replica manager is additionally in charge of removing the unused as well as old replicas. The block is referred to as the fundamental unit of storage in a data center.

3.1 Dynamic Cost-Aware Re-replication and Re-adjusting Strategy

On cloud frameworks, together with the accessibility, fault maintenance, performance, the replication expense also turns to be a critical component. The method [48] proposed an algorithm referred to as Dynamic Cost-Aware Re-replication and Re-adjusting Strategy (DCR2S) towards a heterogeneous cloud computing framework [49-56]. To begin with these proposed algorithms initially it ought to be designed with heterogeneous cloud framework to comprehend the replication cost, availability of the file, and framework availabilities relationship. The introduced algorithm includes three strides and these strides can be utilized to replicate the data files. At the initial stride, the more popular or efficient data file was selected depending upon the intensity of access together with that the decision has been generated to decide the event when the replication function should be invoked [57-65]. At the second stride, the new replicas ought to be established for the preferred data file thus to face the availability needs.

III. RESULTS AND DISCUSSIONS

Here, we have plotted the graph for all the three techniques with the varying number of replicas in x-axis and the corresponding parameters in y-axis. The performance evaluation is carried out with five different parameters namely the SBER (System Byte Effective Rate), Load, Response time, Latency and Fault Ratio.
Figure 2 and 3 delineates the percentage of load balance when the number of replicas is maximized for about 5 nodes. When number of replicas is increased in the node, the maintenance of load in node turns to be difficult.

IV. CONCLUSION

In this comparative study, we have presented an extensive analysis of three different data replication strategies on a cloud environment. The performance study of these three data replication techniques was carried out in terms five different parameters. Here, the analysis has been made with the help of by varying the number of nodes from 5 to 10 with respect to number of replicas. For all the different analysis, graphs were plotted and the detailed analysis was performed to identify the efficient technique from the three different techniques taken for experimental analysis. Finally, EPA [3] when compared with ADRS [1]and DCR2S [2] provides better performance after evaluating it with different parameters.

REFERENCE


Detection using Data Mining Classification Techniques. International Journal of Advanced Research in Computer Science, 3(1).


[39] Le, D.-N.a, Kumar, R.b, Nguyen, G.N., Chatterjee, J.M.d, Cloud Computing and
Virtualization, DOI: 10.1002/9781119488149, Wiley.


In 2015 International Conference on Soft-Computing and Networks Security (ICSNS) (pp. 1-5). IEEE.


