International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 05 | May -2017 www.irjet.net p-ISSN: 2395-0072

Data Replication in Cost Optimized Geo-Distributed Clouds for OSN

Yashavantha kumar G M[1], Srinivasa A H[2]

[1] *M. Tech Student , Department of Computer Science & Engineering, Dr. Ambedkar institute of Technology,Bangalore-560056*
[2] *Associate Professor, Department of Computer Science & Engineering, Dr. Ambedkar institute of Technology,Bangalore-560056*

---***---

Abstract – *To Deploy online social networks services geo-distributed clouds provides good platform. Geo-Distributed clouds having the multiple data centers or servers around the world. The major concern of OSN services to provide a High Data availability by replicating the data, fault tolerance, satisfactory Quality of service(QOS) and increasing the performance ,data replication scheme significantly improve the efficiency of online social networks(OSN's).And this paper is also focused on the monitory cost spent on cloud resources we design the algorithm called cos-play. it will reduce the one-time cost spent on cloud resources when continuously evaluated over a 48 months over OSN and the results shows that this work gives the better data avalibility,Quality of service, data security by replication and minimizes the initial one-time cost when compared to other methods of real world scenarios.*

***Key Words*: Data Replication, Cloud computing, online social networks, Performance analysis, geo-distributed clouds.**

1.INTRODUCTION

Internet applications are progressively necessary to our everyday lives, as we tend to consider them for email, searching, on-line storage, on-line business, and far additional. These applications face an information measurability challenge that's obtaining worse, for 2 reasons. First, there's a growing range of users in AN increasing range of regions. And second, the storage wants per user ar growing as a lot of applications become accessible on-line, users accumulate additional information, and systems collect additional info from users to focus on ads and personalize their expertise[1][2]. As a result, these applications got to be geo-distributed, which implies they're deployed across multiple information centers round the world, as a result of constraints on the scale, bandwidth, and power consumption of one information center. Besides providing data availability, geo-distribution additionally permits a user to be served from a close-by information center[8], thereby reducing user response times and information measure consumption. For that, the user's information ought to be at the correct information center, namely, an information center near the user[4]. this can be referred to as access neighborhood.

clouds offer a very important platform for deploying distributed on-line services. apparently, these 2 changes tend to be combined. whereas OSN services usually have a really massive user base and wish to scale to fulfill demands of users worldwide, geo distributed clouds that offer Infrastructure-as-a-Service will match this want seamlessly and supply tremendous resource and price potency benefits. Infinite on-demand cloud resources will accommodate the surges of user requests; versatile pay as you go[11] charging schemes will save the investments of service suppliers; and cloud infrastructures additionally free service providers[3] from building and operative one's own information centers.

Migrating OSN services toward geographically distributed clouds should reconcile the requirements from many completely different aspects[7]. First, OSN suppliers need to optimize the initial price spent in mistreatment cloud resources. for example, they will want to reduce the storage price once replicating users' information at over one cloud[1][4], or minimize the inter cloud communication price once users at one cloud need to request the information of others that square measure hosted at a distinct cloud. Moreover, OSN suppliers[11][3] hope to supply OSN users with satisfactory quality of service (QoS).

To finish this, they'll need a user's information and people of their friends to be accessible from the cloud nearest to the user[9], as an example. Last however not least, OSN suppliers can also worry with information availableness, e.g., guaranteeing the quantity of users' information replicas to be no fewer than a fixed threshold across clouds. Addressing all such wants of price, QoS, associate degreed information accessibility is more sophisticated by the actual fact that an OSN endlessly experiences dynamics, e.g., new users be part of, previous users leave, and social relations also vary[8].

Existing work on OSN service provisioning either pursues least price in an exceedingly single web site while not the QoS concern as within the geo-distribution case [6], [5] or aims for least inter-data-center traffic within the case of multiple information centers while not considering different dimensions of the service [6], e.g., information accessibility. additional significantly, the models altogether such work don't capture the initial price of resource usage and therefore cannot match the cloud situation. There is some work on cloud-based social video [8],[9], that specialize in investing on-line social relationships to enhance video distribution, that is just one among the various aspects of OSN services; most improvement analysis on multi cloud and multi-data-center services isn't for OSN [4], [3],[12]. They fail to capture the OSN options like social relationships and user interactions, and therefore their models don't seem to be applicable to OSN services.

Deploying additional servers is a intuitive path to build the system scale, except for the most effective performance one has to verify wherever best to place the information, whether or not replication is required, and, if so, how. This paper is concentrated on replication; specifically, we tend to propose S-clone, a socially-aware information replication theme which may considerably improve a social network's potency by taking into consideration social relationships of its information[5]. S-clone's performance is supported in our analysis study. during this paper, we have a tendency to study the matter of optimizing the financial price of the dynamic, multi cloud-based OSN whereas guaranteeing its QoS and information availableness.

2. RELATED WORK

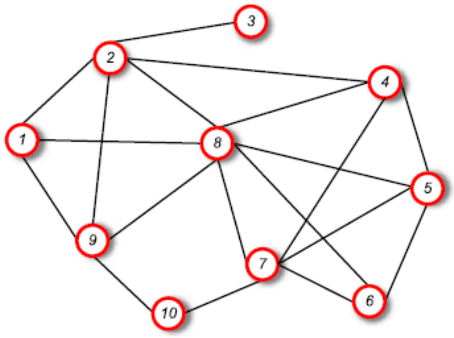
The main approaches to rising an information system's scalability: vertical scaling and horizontal scaling. While vertical scaling scales[9] up the system by adding additional hardware resources to the present servers, horizontal scaling scales "out" the system instead, by adding goods servers and partitioning the work across these servers. Vertical scaling is easy to manage, however horizontal scaling is additional price effective and avoids the single-point of- failure and bottleneck issues. The latter has been a actual commonplace once it involves managing information at large scale for several OSNs nowadays.

We initially model the price, the QoS, and also the information accessibility of the OSN service upon clouds. Our value model identifies differing kinds of prices related to multicloud OSN whereas capturing social neighborhood [6], [5], a very important feature of the OSN service that almost all activities of a user occur between herself and her neighbors. guided by existing analysis on OSN growth and our analysis of real-world OSN dynamics, our model approximates the whole price of OSN over consecutive time periods once the OSN is massive in user population however moderate in growth, sanctioning United States of America to attain the improvement of the full value by severally optimizing the price of every amount. Our QoS model links the QoS with OSN users' knowledge locations among clouds. for each user, all clouds offered area unit sorted in terms of a particular quality metric (e.g., access latency); thus, each user will have the foremost most popular cloud, the second most most popular cloud, etc. The QoS of the OSN service is best if a lot of users have their knowledge hosted on clouds of the next preference. Our knowledge availableness model relates with the minimum range of replicas maintained by every OSN user. supported these models, we have a tendency to then formulate the price improvement downside that considers QoS and knowledge availableness needs.

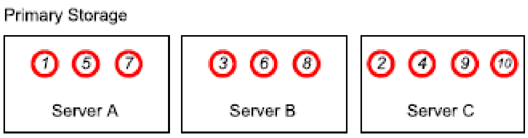
3. RECENT METHODS

S-clone relies on associate intuition that if we want to put a duplicate copy for a user i somewhere, the foremost fascinating location ought to be the first server of most neighbors of i; this manner, most neighbors can have the benefit of this duplicate after they issue a browse question. Below[5], we have a tendency to gift the recursive details of S-clone initial for the static case wherever we want to duplicate the information for a set social graph, so for the dynamic case during which we want to create this replication adapt to changes within the social graph.
The planned replication procedure consists of 2 phases:
1. Replicate section: The procedure starts during this phase. It works during a greedy manner, consecutive considering a node at a time and finding the simplest path to replicate its information. Suppose that the nodes ar processed within the order . for every node i (initially node 1) that has not been processed (i.e., all the nodes 1, 2, i one are processed), the K replicas ar appointed to their corresponding servers as follows. First, the placement bar graph below is computed for i
2. modify section: The goal of this phase is to search out the servers for the remaining replicas that can't be placed as a result of the special case said higher than. as a result of the scan price is that the same despite wherever we have a tendency to store these replicas, their locations are chosen to maximise load equalisation. the most effective path to do that is to method every remaining duplicate one by one and place it on the server that presently has less amount load.After the K replicas of i is placed been placed, the algorithmic program can method future user, node (i + 1). Note that in selecting the K servers to duplicate the information for a user, we tend to don't take into account its primary server as a result of it makes no sense to place a duplicate and its primary on a similar server. Also, as solely the placement data of the first copies is employed to see wherever to duplicate a user, the order to method the users doesn't have impact on the ultimate duplicate placement. this {can be} a fascinating property as a result of there is also totally different structures to represent a social graph and S-clone can work with any structure unbiased.

With an example, we tend to compare S-clone[5] to the random replication approach once they are applied to a 10-node social graph shown in Fig. 1a. we tend to assume AN existing partition of users across 3 servers, A, B, and C (Fig. 1b) and wish to own one duplicate for every user (K = 1). Fig. 2a shows the results of employing a random replication algorithmic program on high of this partition, whereas Fig. 2b shows the results of running S-clone during which the order of visiting the users is random (4–1–3–2–10–5–9–6–8–7). beginning with user four, server A is that the primary server for 2 neighbor users of four and server B is that the primary server for less than one neighbor. Therefore, server A is chosen to duplicate user four. The remaining users ar replicated equally. within the case of user eight, servers A and C every function primary server for 3 neighbors of eight, however to interrupt the tie, server C is chosen to duplicate eight as a result of it's up to now keep fewer copies (6 copies, compared to seven of server A). For an analogous reason, user seven is replicated on Server B however not server C. Table one summarizes the price to scan the information for every of the 10 users, showing an apparent improvement of S-clone over random replication (24% better).

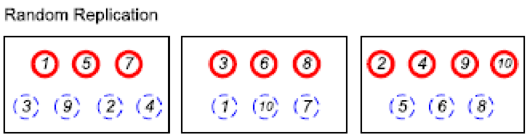


(a) Social graph

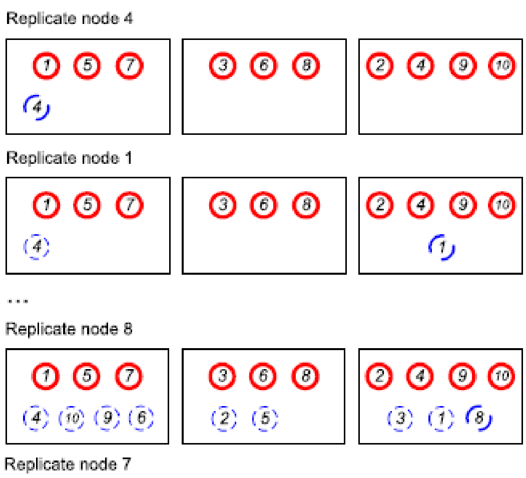


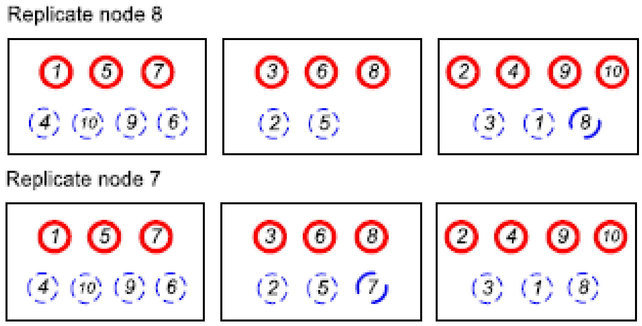
(b) Primary partition

Fig 1. Given social graph and its primary partition ($M = 3, K = 1$).



(a) Random replication





(b) S-CLONE

Fig 2. S-CLONE vs. random replication.

4. PROPOSED WORK

In this paper, we are going to study dificulties of optimizing the initial price of the dynamic, multi cloud-based OSN whereas guaranteeing its QoS and information availableness. We initial model the price, the QoS, and also the information availableness of the OSN service upon clouds. Our price model identifies differing types of prices related to multi cloud OSN whereas capturing social neighbourhood [6], [5], a very important feature of the OSN service that almost all activities of a user occur between herself and her neighbors. guided by existing analysis on OSN growth and our analysis of real-world OSN dynamics, our model approximates the entire price of OSN over consecutive time periods once the OSN is massive in user population however moderate in growth, sanctioning USA to realize the improvement of the overall price by severally optimizing the price of every amount[10]. Our QoS model links the QoS with OSN users' information locations among clouds.For every user, all clouds accessible ar sorted in terms of an exact quality metric (e.g., access latency); thus, each user will have the foremost most popular cloud, the second most most well-liked cloud, etc. The QoS of the OSN service is best if additional users have their information hosted on clouds of a better preference. Our information availableness model relates with the minimum range of replicas maintained by every OSN user.

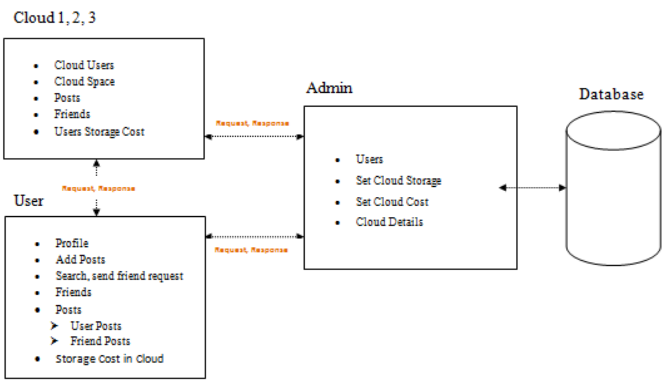


Fig-3 Architecture Diagram

1) A user includes a storage price, that is the initial price for storing one duplicate of her information[10] (e.g., profile, statuses) within the cloud for one charging amount.
2) same as, a user includes a traffic price, that is the financial price throughout a charge amount as a result of the intercloud traffic. As mentioned earlier, as a result of social neighborhood, in our settings the intercloud traffic solely involves writes (e.g., posting tweets, leaving comments). we have to don't think about intracloud traffic, regardless of browse or write, because it is free from charge.
3) A user includes a sorted list of clouds for the aim of QoS.

4.1 Algorithm: cos-play

This work uses a series of role-swaps to maximize the overall price reduction whereas information accessibility and guaranteeing QoS needs. Our algorithmic rule follows a greedy approach in exploitation role-swaps and requiring that each applied role-swap decrease price. The additional price reduction every role-swap has and also the additional role-swaps ar applied, the total price reduction we are able to accomplish. Note that our algorithmic rule computes a much better placement, and it doesn't physically manipulate information[10]. once our algorithmic rule terminates, information ar role-swapped or touched (in the case of redistribution) from existing locations to new locations so as to implement the new placement output by our algorithmic rule. we tend to describe our cos-play algorithmic rule as follows: initial with AN existing placement, the formula runs and repeats two procedures called single role swaps and double role swaps.

5. RESULTS

The paper results show that real world social media data over multiple clouds it gives the results about user profiles, information about cloud storage and we can set the cloud cost as per the requirement, total friends in the cloud and different cloud locations in the different geo graphical regions. The algorithm can reduce the initial cost of the cloud resources and increasing the data-availability.

5.1 Set Cloud Storage

In this module Fig. 4 we can set the cloud storage and this module has sub modules called clouds and set space.

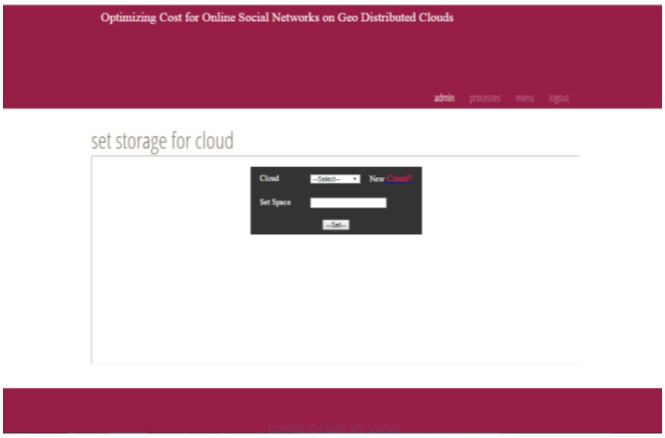


Fig -4: Set Cloud Storage

First we can select the cloud based on the location and Second we can also set the storage space for the cloud provided by the service provider.

5.2 Set cloud cost

This module Fig. 5 gives the information about the cloud cost based on the cloud storage usage.

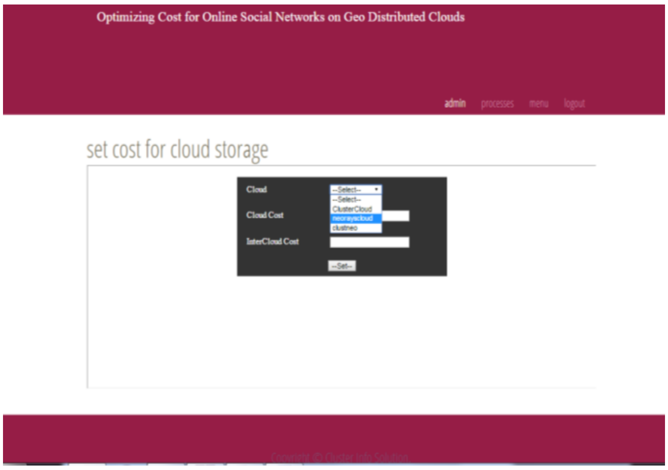


Fig -5: Set Cloud Cost

Inter-cloud charges is also apply if the user trying to access the another cloud.

5.3 Cloud Details

This module gives the information about clouds, locations of the cloud, Total memory used, Remaining memory, cloud cost, inter cloud cost as shown in the Fig.6.

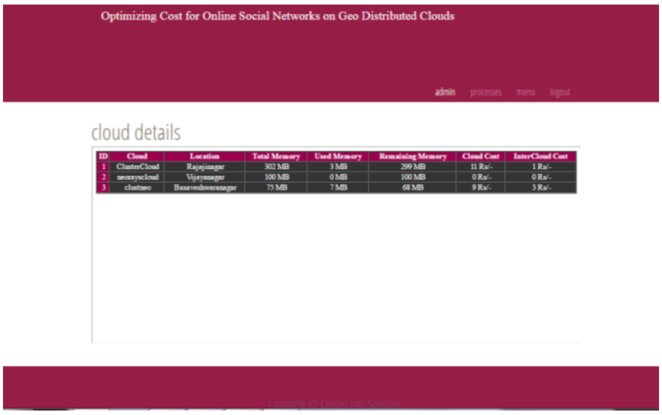


Fig -6: Cloud Details

5.4 Cloud Users

This module gives the information about total users are present in the cloud with their Image, Name, Location and Contact Details etc .as shown in Fig.7

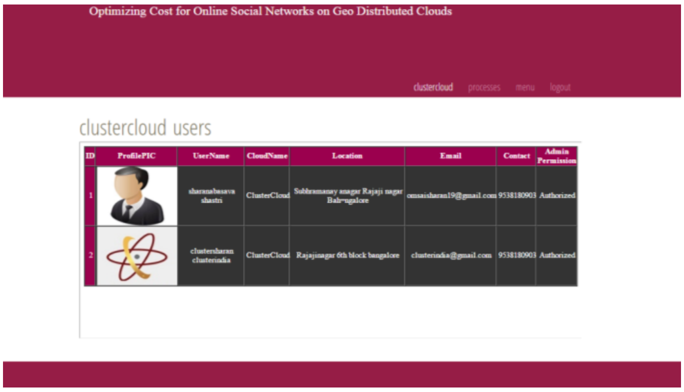


Fig -7: Users of Cloud

And also this work shows the results of different modules like Friends, User Profiles, Posts added in the cloud, Space used by the Users, add posts, View Posts, details of all Friends and their Storage Details.

6. CONCLUSIONS

In this paper, understanding and finding the solution for optimizing the initial cost spent on the cloud resources. And replicating the OSN data to provide the data availability in case of failures in the server and faster the read and write operations, In this paper we used the replication method called s-clone, it is very good method for load balancing and writing the loads across the servers. S-clone is only applicable to the OSN for the data availability ensuring the minimum number of replicas for each user. And ensuring the QOS and data availability by the proposed algorithm **cos-play** resulting in the large scale twitter data and this algorithm is useful for cost reductions of OSN services and reduce the onetime cost and cost reductions over 48 months of service and provide the QOS and data availability as per the user requirements.

REFERENCES

[1] G. Liu, H. Shen, and H. Chandler, “Selective data replication for on-line social networks with distributed datacenters,” in Proc. IEEE ICNP,2013, pp. 1–10.
[2] A. E. Mislove, “Online social networks: measuring, analysis, and applications to distributed data systems,” Ph.D. , Rice University, Houston, TX, USA, 2009.
[3] H. Roh, C. Jung, W. Lee, and D.-Z. Du, “Resource pricing game in geo distributed clouds,” in Proc. IEEE INFOCOM, 2013, pp. 1519–1527.
[4] Y. Sovran, R. Power, M. K. Aguilera, and J. Li, “Transactional storage for geo-replicated systems,” in Proc. SOSP, 2011, pp. 385–400.
[5] D. A. Tran, K. Nguyen, and C. Pham, “S-clone: Socially-aware knowledge replication for social networks,” Comput. Netw., vol. 56, no. 7, pp. 2001–2013, 2012.
[6] J. M. Pujol et al., “The very little engine(s) that could: Scaling on-line social networks,” IEEE/ACM Trans. Netw., vol. 20, no. 4, pp. 1162–1175, Aug. 2012..
[7] N. Tran, M. K. Aguilera, and M. Balakrishnan, “Online migration for geo-distributed storage systems,” in Proc. USENIX ATC, 2011, p. 15.
[8] Z. Wang et al., “Propagation-based social-aware replication for social video contents,” in Proc. ACM transmission, 2012, pp. 29–38.
[9] Y. Wu, C. Wu, B. Li, L. Zhang, Z. Li, and F. C. M. Lau, “Scaling social media applications into geo-distributed clouds,” in Proc. IEEE INFOCOM, 2012, pp. 684–692.
[10] L. Jiao, J. Li, T. Xu, and X. Fu, “ Optimizing cost for on-line social networks on geo-distributed clouds,” in Proc. IEEE ICNP, 2016.
[11] Amazon, Seattle, WA, USA, “Amazon EC2 rating,” [Online].
[12] H. Xu and B. Li, “Joint request mapping and response routing for geo-distributed cloud services,” in Proc. IEEE INFOCOM, 2013, pp. 854–862.