www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 10 October 2017, Page No. 22817-22822
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i10.19

A Novel Method for Data Hosting and Load Balancing in Multi Cloud Environment

S. Jyoti Reddy, Sk. Sofia

M Tech (CS)
Asst. Professor
Priyadarshini College of Engineering and Technology Nellore, Andhra Pradesh

Abstract
The increasing popularity of cloud computing attracted so many organizations to host cloud services. Due to the enormous growth of cloud data centers the users are now able to put their data in any of the cloud data centers. It is difficult for the cloud service provider to choose an appropriate data center based on the user requirements. To ensure reliability cloud service providers uses replicas. It is essential to load balance among the cloud data centers for better performance. We propose a new method for the distribution of data among the cloud data centers based on the Predictor-Distribute Algorithm (PREDA).
Index Terms: *Multi Cloud, Erasure and Replication, Predictor.*

I Introduction.

Cloud load balancing is the process of distributing workloads and computing resources in a cloud computing environment. Load balancing allows enterprises to manage application or workload demands by allocating resources among multiple computers [6], networks or servers. Cloud load balancing involves hosting the distribution of workload traffic and demands that reside over the Internet.
There are various cloud vendors exhibiting variations in working performances and pricing policies. They design with different system architectures and apply various techniques to provide better services. So that customers are unable to understand which clouds are suitable to host their data? This is called vendor lock in risk. It is inefficient for an organization to host all the data in a single cloud. It does not provide guaranteed availability.
Recently, multi-cloud data hosting has received wide attention from researchers, customers, and startups. The basic principle of multi-cloud (data hosting) is to distribute data across multiple clouds to gain enhanced redundancy and prevent the vendor lock-in risk[1], as shown in Fig. 1. The “proxy” component plays a key role by redirecting requests from client applications and coordinating data distribution among multiple clouds.

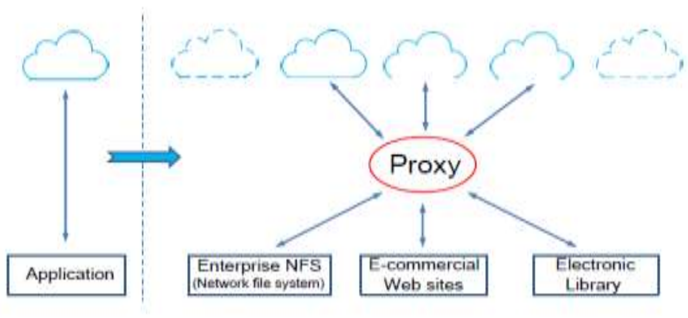


Fig 1

II Related Work.

Qian Zhao, Guangsheng Feng in their work "A Mixed-Prediction based Method for Allocating Cloud Computing Resources" [2-] proposed MPBRA (Mixed Prediction Based Resource Allocation) method to allocate cloud platform resource according to the periodic and aperiodic service requirement. This method first employs FFT (Fast *Fourier Transform)* theory to judge the periodic cycle and uses Markov process to predict aperiodic service resource requirement. This method obtains higher resource utilization efficiency than other traditional ones, and the number of violating SLA is deducted obviously.

Quanlu Zhang, Shenglong in their work proposed [1] a cost-efficient data hosting scheme with high availability in heterogeneous multi-cloud, named “CHARM”. It intelligently puts data into multiple clouds with minimized monetary cost and guaranteed availability. They combined the two widely used redundancy mechanisms, i.e., replication and erasure coding, into a uniform model to meet the required availability in the presence of different data access patterns.

In existing industrial data hosting systems, data availability and reliability is usually guaranteed by ‘replication’ or ‘erasure coding’. In the multi-cloud scenario, we also use them to meet different availability requirements, but the implementation is different. For replication, replicas are put into several clouds, and only read access is served by the “cheapest” cloud that charges minimal for out-going bandwidth and GET operation. For erasure coding, data is encoded into n blocks including n*m coding blocks, and these blocks are put into n different clouds. In this case, though data availability can be guaranteed with lower storage space (compared with replication), a read access has to be served by multiple clouds that store the corresponding data blocks. Consequently, erasure coding cannot make full use of the cheapest cloud as what replication does. Still worse, this shortcoming will be amplified in the multi-cloud scenario where bandwidth is generally (much) more expensive than storage space.

<u>System Architecture</u>

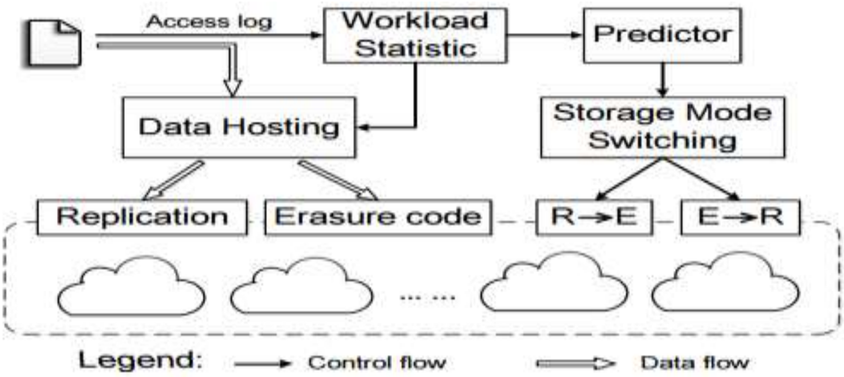


Fig 2

The system performs the following

- Workload Statistic keeps collecting and tackling access logs to guide the placement of data. It also sends statistic information to Predictor which guides the action of SMS.
- Data Hosting stores data using replication or erasure coding, according to the size and access frequency of the data.
- SMS decides whether the storage mode of certain data should be changed from replication to erasure coding or in reverse, according to the output of Predictor.

  However these systems has some pitfalls like, No guaranteed allocation to exact server, more time in allocation of cloud data center and Not Efficient in the point of cost based allocation.

III Proposed Method.

We propose a novel method for cost-efficient data hosting scheme with high availability in heterogeneous multi-cloud based on a predictor model. It intelligently puts data into multiple clouds with minimized monetary cost and guaranteed availability. Specifically, we combine the two widely used redundancy mechanisms, i.e., replication and erasure coding, into a uniform model to meet the required availability in the presence of different data access patterns. Next, we design an efficient Predictor algorithm to choose proper data storage modes involving both clouds and redundancy mechanisms (ERREPLCA).

In existing system the major focus is combining the replication and erasure methods they don't provide a specific method for predictor. However there are many prediction algorithms exists such as weighted moving average method. Some methods use building a classifier to predict the access frequency of files. In the proposed method we build a predictor using data mining algorithms. Since many of the data centres generate enormous log files the size of input is huge we need an algorithm to handle such data.

Advantages:

1. Uses Replication mechanism when the file's size is small. Otherwise uses erasure coding.
2. The storage mode table only depends on prices of the available clouds and required availability. If the prices change, the table will change accordingly.
3. Uses an Efficient predictor to decide the storage mode and a suitable cloud data center.
4. Saves monetary costs.

We use a split algorithm for predicting a data centre with fewer loads for next allocation. Once a Data Centre is allocated later based on statistics given by predictor we apply ERREPLICA method.

Stage1

// Predictor

MakeTree(Training Data *T*)

Partition(*T*)

Partition(Data *S*)

if(all points in S are in the same class)

then return;

Evaluate Splits for each attribute A;

Use best split to partition S into S1 and S2;

Partition(S1);

Partition(S2);

or each attribute *A* **do**

traverse attribute list of *A*

for each value *v* in the attribute list **do**

find the corresponding entry in the class

list, and hence the corresponding class

and the leaf node *l*

update the class histogram in the leaf *l*

if *A* is a numeric attribute **then**

compute splitting index for test ($A \leq v$) for *l*

if *A* is a categorical attribute **then**

for each leaf of the tree **do**

find subset of *A* with best split

// Stage2 Distribute

// choosing a Storage Mode and allocating to a best Data Centre.

The Algorithm

Setup (n datacenters)

Alloc(m) //Allocate m blocks to each dc

Compute Load for each Data Center.

Choose a datacenter based on load

For k=1 to n

    Check the availability of k^{th} dc suitable for μ

If μ = sflag

    Allocate to K

Else

    Ealloc (n, μ)

End

// Algorithm for partitioning and choosing a suitable cloud with least cost.

Ealloc (n,μ)

//The output is minimum cost C, The set of the selected clouds H.

1.C←inf;

2.H={} //initially empty.

3.Sort the clouds by S+ μ // Accessibility

4. for m= 1 to n do

    A←calculate the availability of G

    If A<=Amax then

        Mcost←minimamal cost.

    If Mcost<C then

        H←G.

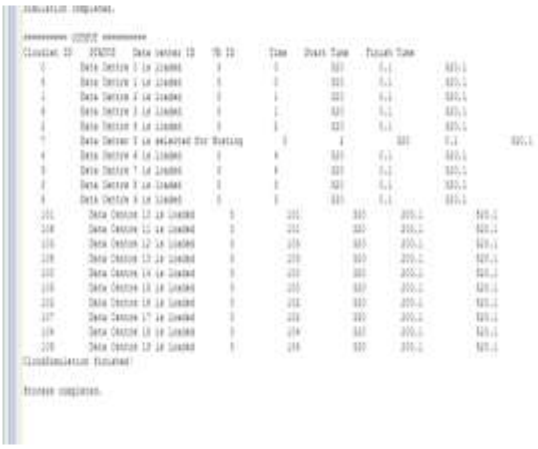
    End

IV Experimental Setup

The proposed System was developed in Java with cloud simulator3.0. Results are tested using VMWare player 12.5.7 on a 64 bit machine .The systems is tested for 10 Data centers which are set using simulator.

Fig3: Simple Simulation.

Fig4: Uploading of a File to a Data Centre.



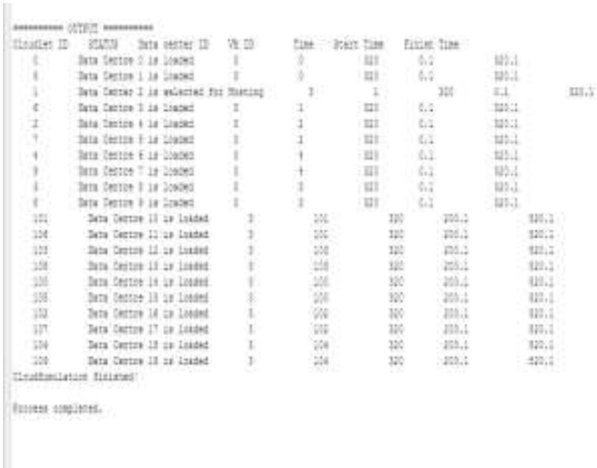


Fig5: Data centre with fewer loads (5) chosen

Fig6: Data center 2 with Less Load for Hosting

V Conclusion:

In this paper we proposed a method for data hosting in cloud environment. This method uses a predictor which determines the suitable data centre for cost efficient allocation. The predictor uses Data mining split algorithm for prediction. We also used erasure and replication code for secure storage.

References

1. Quanlu Zhang∗ , Shenglong Li " A Cost-efficient Multi-cloud Data Hosting Scheme with High Availability.", IEEE Transactions on Cloud Computing VOL. X, NO. XX, 2015.
2. Qian Zhao, Guangsheng Feng "A Mixed-Prediction based Method for Allocating Cloud Computing Resources ", International Journal of Grid Distribution Computing Vol.8, No.2 , pp.201-212 .(2015)

3. F.Salfner “A survey of online Prediction Methods “ ACM computer survey vol 42,No 3 pp 10:1-10:42 (2010)
4. S. S. MOHARANA, R. D. RAMESH, D. POWAR, “Analysis of load balancers in cloud computing” International Journal of Computer Science and Engineering,2013,vol 2,pages 101-108
5. S. Liu, X. Huang, H. Fu, and G. Yang, “Understanding Data Characteristics and Access Patterns in a Cloud Storage System,” in CCGrid. IEEE, 2013.
6. Sahu, Yatendra and Pateriya, RK, “Cloud Computing Overview with Load Balancing Techniques”, International Journal of Computer Applications, 2013,vol. 65, Sahu2013
7. Chaudhari, Anand and Kapadia, Anushka, “ Load Balancing Algorithm for Azure Virtualization with Specialized VM”, 2013,algorithms,vol 1,pages 2, Chaudhari
8. Nayandeep Sran,Navdeep Kaur , “Comparative Analysis of Existing Load Balancing Techniques in Cloud Computing ”,vol 2,jan 2013
9. Bala, Anju and Chana, Inderveer, “A survey of various workflow scheduling algorithms in cloud environment”, 2nd National Conference on Information and Communication Technology (NCICT), 2011.
10. D. Bhu Lakshmi1 and S. Arundathi, “Providing Privacy and Security for Cloud Data Using Data Mining,” International Journal of Innovation and Scientific Research Vol. 11 No. 2, Nov 2014.
11. A. Bhadani and S. Chaudhary, “ performance evaluation of web servers using central load balancing policy over virtual machine on cloud”, proceedings of third Annual ACM.
12. J. M. Galloway, K. L. Smith, and S. S. Vrbsky, “Power aware load balancing for cloud computing,” in Proceedings of the World Congress on Engineering and Computer Science, vol. 1, pp.19–21, 2011.
13. S. Sethi, A. Sahu, and S. K. Jena, “Efficient load balancing in cloud computing using fuzzy logic,” IOSR Journal of Engineering, vol. 2, no. 7, pp.65–71, 2012.
14. Z. Nine, M. SQ, M. Azad, A. Kalam, S. Abdullah and R. M. Rahman, “fuzzy logic based dynamic load balancing in virtualized data centers” In fuzzy system (FUZZ), IEEE International conference on, pp. 1-7, 2013.